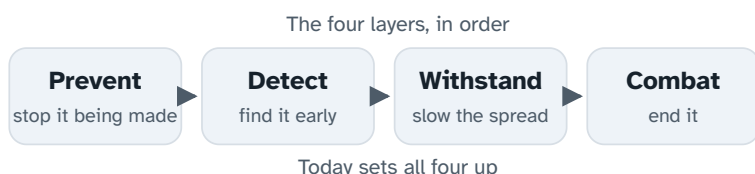


Day 1: Building a Pandemic-Proof World

Monday 26 October · discussion 3:00 to 4:50 pm New York

THE DAY MAP

How bad have pandemics been, how are AI and synthetic biology changing the threat, and what would a pandemic-proof world actually look like?



19 readings here, about 145 minutes of reading. Discussion Monday 26 October, 3:00 to 4:50 pm New York, with Conor Spence.

1.2-1 What were the death tolls from pandemics in history? 2 + 10 min

[original paper](#) →

1.2-2 An illustrated history of the world's deadliest epidemics, from ancient Rome to Covid-19 3 + 0 min

[original paper](#) →

<u>1.2-3</u>	A Short History of Biological Warfare: From Pre-History to the 21st Century original paper →	3 + 0 min
<u>1.2-4</u>	Germ Warfare: A Very Graphic History original paper →	3 + 11 min
<u>1.2-5</u>	Historical Overview: From poisoned darts to pan-hazard preparedness original paper →	3 + 0 min
<u>1.2-6</u>	High-risk human-caused pathogen exposure events from 1975-2016 original paper →	3 + 0 min
<u>1.3-1</u>	AI Is Reviving Fears Around Bioterrorism. What's the Real Risk? original paper →	3 + 0 min
<u>1.3-2</u>	Mother Nature is not 'the ultimate bioterrorist' original paper →	3 + 7 min
<u>1.3-3</u>	AI and the Evolution of Biological National Security Risks original paper →	3 + 0 min
<u>1.3-4</u>	Nature Sets a Baseline, Not a Ceiling, for Biological Harm original paper →	3 + 8 min
<u>1.3-5</u>	How AI could enable catastrophic pandemics original paper →	3 + 9 min

<u>1.3-6</u>	CBRN evaluations, Claude Opus 4 System Card original paper →	3 + 0 min
<u>1.3-7</u>	The New Bioweapons: How Synthetic Biology Could Destabilize the World original paper →	3 + 13 min
<u>1.3-8</u>	Securing Civilisation Against Catastrophic Pandemics original paper →	3 + 7 min
<u>1.3-9</u>	Mirror Life: Brief of the Scientific Advisory Board original paper →	3 + 11 min
<u>1.4-1</u>	Pandemic prevention as fire-fighting original paper →	3 + 13 min
<u>1.6-1</u>	Biosecurity Really: A Strategy for Victory original paper →	3 + 0 min
<u>1.6-2</u>	Global Catastrophic Biological Risks A Guide for Philanthropists original paper →	3 + 0 min
<u>1.6-3</u>	Securing Civilisation Against Catastrophic Pandemics original paper →	3 + 0 min

WHAT THE COURSE ASKS YOU TO DO

1.4 Envisaging a pandemic-proof world

Imagine the year is 2050 and the world is much more pandemic-proof than it is today. What kinds of technologies, policies and organisations exist, and how might we get there?

We encourage you to use your existing expertise, as well as taking inspiration from the ‘Pandemic Prevention as Fire-Fighting’ article or other research.

This is only a starting point and we don’t expect comprehensive answers – use this as a brainstorming exercise to think of a broad range of ideas. We’ll dig deeper into some of these interventions later in the course.

Things you might want to consider:

What could a day in your life look like? - What technologies might you interact with that prevent you catching or transmitting pathogens?

What could we copy from emergency preparedness and response in other settings (fire, war, cyber etc.)? - Are there analogous technologies, policies or organisations in those areas that we could build in the biosecurity space?

Consider grouping your ideas into outbreak prevention, detection and response (note these categories aren’t always clear-cut!).

THE READINGS

What were the death tolls from pandemics in history?

Saloni Dattani · 2023

[Read the original paper →](#) · the course assigns: whole piece, including the appendix on major pandemics

Pandemics have killed on a scale almost nothing else in history matches, and the only reason we can say so is that researchers rebuilt the death tolls from tax registers, parish records, archaeology and excess-mortality models, each with its own uncertainty.

Average summary

Dattani's argument is that we **underestimate** how often and how badly pandemics have hit us, because our ancestors could not count their own dead.

Records exist for some countries and not others. Often nobody knew which pathogen was involved, so we cannot always say whether one event was a pandemic or several outbreaks at once.

So the piece is as much about method as about history. Dattani lays out the four ways researchers put numbers on old pandemics. Excess deaths: how many more people died than would have died in a normal year. Net population reduction: the difference in population size before and after, used for severe

events like the Columbian Exchange. Epidemiological modelling: working from transmission, geographic spread, fatality rate and access to treatment. And recorded (confirmed) deaths, which she says may vastly underestimate the real toll, since cause-of-death registration is still missing in much of the world.

Then the numbers. The Black Death of 1346 to 1353 killed around 50 to 60% of Europe's population, roughly 50 million people, in six years. *Yersinia pestis* also caused the pandemic that began in 541 CE and the third pandemic of 1894 to 1940. The Columbian Exchange after 1492 shows an estimated 48 million deaths: a Native American population of about 54 million before Columbus fell to 5.6 million by 1600, a reduction of about 90%. The 1918 influenza pandemic killed an estimated 50 to 100 million. Seven cholera pandemics have occurred in the last two centuries; in India alone, at least 23 million people died of cholera between 1865 and 1947. HIV/AIDS has killed an estimated 33 million people between 1981 and 2022. COVID-19 shows around 27 million excess deaths between January 2020 and November 2023.

Her closing point is hopeful and conditional. Before germ theory we did not know what caused these diseases or how to protect ourselves. Progress came not only from science but from data collection, public

health, healthcare access and cooperation: death records let John Snow's successors work out how cholera spread, coordination on HIV/AIDS prevented millions of deaths, global testing for new influenza strains lets the flu vaccine be adapted every year. With better understanding and effort the world can respond faster. Without it, she writes, we will keep facing major pandemics as we have so far.

Why it's assigned

This is the first reading of the course and it supplies the baseline for the day's first question: how bad have pandemics actually been? The course's own note says Our World in Data traces how pandemics from the Black Death to COVID-19 have repeatedly devastated humanity. It also teaches, without making a fuss about it, the habit the rest of the week depends on: asking where a number came from before you use it. Unit 1 sits before the four defensive layers (prevent, detect, withstand, combat); this reading is the "how big is the problem" half of it.

In depth

The problem with the historical record

Dattani opens by saying COVID-19 pushed pandemics into public consciousness, but they have afflicted humanity for millennia: influenza, cholera, bubonic plague, smallpox, measles. Our ancestors were largely **powerless** against these diseases and unable to

evaluate their true toll. Without good record-keeping of cases and deaths, the impact of outbreaks was underrecognised or even forgotten. The result, she says, is that we tend to underestimate both the frequency and the severity of pandemics in history.

Two specific gaps: we often have records from some countries and not from others, even when we know the disease's geographic reach was very wide; and we often do not know which pathogens caused an outbreak, so we cannot say whether an event was one pandemic or parallel outbreaks of different diseases.

What counts as a pandemic

There is no universally accepted definition. Diseases called pandemics share some features: a vast geographic range (a continent or several), rapid growth or expansion in range, high infectiousness, large numbers of people affected, and a novel pathogen against which there is little or no pre-existing immunity.

In the endnotes she adds that the World Health Organization defines pandemics only in the context of influenza: a novel influenza virus emerges, it can sustain human-to-human transmission, it causes disease, and the human population has little or no immunity. For diseases in general there is no official WHO definition of a pandemic. The highest outbreak response the WHO designates is a Public Health

Emergency of International Concern, a formal declaration for an extraordinary event that constitutes a public health risk to other states through international spread and may require a coordinated international response. A PHEIC carries a legal obligation for countries to respond promptly.

The four estimation methods

Excess deaths: the additional number of deaths during a pandemic compared with the expected number in a typical year. Useful when death certificates are unavailable.

Net population reduction: the difference in population size before and after. Used for severe events where a significant fraction of the population died, such as the Columbian Exchange.

Epidemiological modelling: built from knowledge of transmission, geographic spread, the fatality rate (the share of people affected who die), access to treatment and other data.

Recorded or confirmed deaths: deaths officially reported with the disease as the cause. This method may vastly underestimate the toll, because comprehensive historical records are lacking, and even today cause-of-death registration is missing in many parts of the world. That is why confirmed COVID-19 deaths are far below the total.

The figure

The article's central chart is a timeline. Each pandemic is a circle whose size represents its estimated death toll; pandemics with no known death toll appear as triangles instead. What to notice: the largest circles, such as the Black Death, represent pandemics that killed more than half the population; and some diseases recur in waves, with seven major cholera pandemics and another seven caused by influenza in just the last two hundred years.

The Black Death

The Black Death (1346 to 1353 CE) is one of the earliest pandemics with a methodically estimated death toll: around 50 to 60% of Europe's population, approximately 50 million people, in just six years. Researchers know large outbreaks also occurred in Western Asia, the Middle East and North Africa, but no comprehensive global estimate is available.

There were no population censuses, so the European estimate comes from tax and rent registers, parish records and archaeological remains. Uncertainty remains: the records come from a limited number of European regions and are extrapolated to the continent using demographic estimates. Careful examination has led historians to revise the toll upwards and confirmed the cause as the bacterium *Yersinia pestis*.

Her endnote lays out the disagreement honestly. Benedictow estimates upwards of 60% of Europe died; Aberth estimates 51 to 58%; Christakos and colleagues put it at 40%; Jedwab, Johnson and Koyama estimate 39% of Western Europe. Dattani gives a rounded 50 to 60% using Aberth and Benedictow. She also explains why twentieth-century estimates were too low: earlier records came disproportionately from the nobility, who had better nutrition and hygiene and likely lower fatality rates. And she names two remaining problems: the method measures the net change in mortality, so it can include other events of the period and does not account for declines in other causes of death; and the available records are limited to parts of Spain, Italy, France and England.

People in the fourteenth century did not know about this bacterium, nor that it was transmitted from rat fleas to humans, long before germ theory arrived in the late nineteenth century. Without that knowledge they had little idea how to protect themselves. Even after the first wave, the pandemic continued with frequent smaller outbreaks until around 1690. Y. pestis caused the diseases known as bubonic and pneumonic plague: fevers, chills, vomiting, excruciating headaches, and the distinctive "buboes" in swollen lymph nodes in the groin, thighs, armpits or neck. As the bacterium spreads through the lymph nodes it emits toxins that break down blood vessels

and form clots, potentially blocking blood circulation and leading to death.

The other plague pandemics

Genetic evidence suggests *Y. pestis* emerged at least 4,000 years ago. The first known bubonic plague pandemic began in 541 CE with recurrent outbreaks until the mid-eighth century, affecting the Eastern Roman (Byzantine) Empire, the Middle East, North Africa and the Mediterranean; its initial and most severe outbreak, the Justinianic Plague of 541 to 549 CE, is named after Emperor Justinian, who then ruled Constantinople. The third pandemic occurred between 1894 and 1940, mainly in Asia and Africa. Bubonic plague is less common today thanks to improved sanitation and hygiene that reduce rat and flea density, better public health surveillance and effective antibiotics, but cases still appear on different continents, mainly in small towns and villages.

The Columbian Exchange

The chart shows an estimated 48 million deaths. The Columbian Exchange is the period after Columbus's 1492 voyage, when populations, ideas and crops such as tomatoes, potatoes and maize moved between the Americas and the rest of the world. It also involved extensive war, conquest, slavery and the spread of multiple deadly diseases that devastated indigenous populations. Smallpox, cholera, measles, diphtheria,

influenza, typhoid fever, bubonic plague and others had already killed many in Europe, but were more severe for Native Americans, who had been isolated from them and lacked immunity.

The toll is calculated as a net population reduction against the pre-1492 population. The Native American population was estimated at around 54 million before Columbus arrived; over the following century around 48 million died and the population fell to 5.6 million in 1600, a reduction of about 90%. Both figures come from archaeological records, tribal records, censuses, epidemiological modelling, and land and crop use. Dattani flags in her endnote that pre-Columbian population estimates are generally uncertain, partly because disease may have spread faster and further than Europeans travelled, and that recent archaeological work on land use has revised the initial population upwards.

Influenza

Flu pandemics arise through sudden evolutionary changes when different strains combine into novel strains (a process called antigenic shift) that can be more infectious and lethal than previous ones. Flu has affected humanity for thousands of years, but comprehensive death tolls exist only for the pandemics of the last 140 years.

The largest, the 1918 "Spanish flu", has an estimated death toll of 50 to 100 million, compiled from recorded death tolls and regional excess-death estimates. The chart also shows the 1889 "Russian flu" (an estimated 4 million deaths), the 1957 "Asian flu" (2 million), the 1968 "Hong Kong flu" (2 million) and the 2009 "Swine flu" (100,000 to 1.9 million). These come from excess mortality compared with the years immediately before and after, using national mortality records extrapolated to the global population. She notes that the names of flu pandemics describe where they were first reported, not necessarily where they began. Her endnote records that Johnson and Mueller arrived at 50 million and stated the true toll may be up to 100 million; earlier historians ranged from 30 million globally to 18 million in India alone; and a 2018 reassessment estimated around 17 million, which she thinks is likely underestimated because poorer countries saw higher flu mortality rates.

Cholera

Seven cholera pandemics have occurred in the last two centuries. Most are considered to have originated in the Indian subcontinent and spread across countries and continents through war, travel and international trade. Global knowledge of the total toll is limited, but between 1865 and 1947 at least 23 million people died of cholera in India alone, and

significant outbreaks were recorded in many more countries.

Cholera is particularly severe because, untreated, *Vibrio cholerae* can cause severe dehydration and death within hours or days of the first symptoms. Its severity has been reduced by a chain of scientific advances: the understanding that it spreads through contaminated water and food, and therefore that clean water and sanitation prevent it; the identification of *Vibrio cholerae* as the cause; antibiotics; and the knowledge that simple rehydration treatment substantially reduces severe disease. It still kills: since 1960, over 900,000 deaths have been recorded globally as part of the seventh cholera pandemic. An endnote adds that during nineteenth-century outbreaks, between 30 and 60% of people diagnosed with cholera died in many major cities including New York and Berlin; today the case-fatality rate ranges between 0 and 10% globally.

HIV/AIDS

When HIV/AIDS was first identified in the early 1980s it had a fatality rate of 100%, and patients had a median survival time of about one year after diagnosis. HIV attacks white blood cells, which are critical to immune function, leaving patients open to a wide range of opportunistic infections and diseases.

The timeline shows an estimated 33 million deaths worldwide between 1981 and 2022, from available data plus statistical modelling that considers transmission characteristics, behavioural and clinical data, treatment availability and recorded deaths from countries with high death registration. The global response involved international cooperation, resource allocation and antiretroviral therapy, which turned HIV from a fatal diagnosis into a manageable chronic condition with treatment. In recent years around 1.5 million deaths have been averted annually because of antiretroviral therapy, which prevents the virus replicating and so reduces both severity and spread; around 600,000 deaths still occur annually.

COVID-19

COVID-19 was caused by the novel coronavirus SARS-CoV-2, which emerged at the end of 2019. The timeline shows around 27 million excess deaths between January 2020 and November 2023, making it one of the deadliest pandemics of the last century.

Dattani is explicit about why she uses excess mortality: global confirmed deaths are certainly much lower than the real total, because testing was very limited in many countries throughout the pandemic and cause-of-death registration is still lacking in many. Excess mortality also captures deaths indirectly caused by the pandemic's impact on healthcare systems and economies, not only deaths from the

virus. Researchers use national mortality data where available and statistical models elsewhere, drawing on testing rates, confirmed cases and deaths, population age structure and state policies.

She explains her source choice. The 27 million figure comes from The Economist, chosen because the estimate is continuously updated and the methodology is well documented. The WHO and the Institute for Health Metrics and Evaluation have also estimated excess deaths, but their latest estimates only ran to the end of 2021. For that period all three are similar: 18.2 million (IHME), 14.8 million (WHO) and 17.8 million (The Economist). The final chart shows excess mortality alongside confirmed deaths, and she points out the wide uncertainty band around excess deaths while noting that even the lowest estimates are much higher than the confirmed count.

Words to know

- **Pandemic** — a disease outbreak across a continent or several, growing fast, highly infectious, affecting many people, usually caused by a pathogen we have little immunity to. There is no official definition for diseases in general, which is why the boundary cases in this article are argued over.
- **Excess deaths (excess mortality)** — how many more people died in a period than would have been expected in a normal year. It is the main tool for measuring pandemics where testing and death

registration are weak, and it captures indirect deaths too.

- **Net population reduction** — the population before an event minus the population after. Used when a pandemic was so severe that the population drop itself is the measurement, as with the Columbian Exchange.
- **Case-fatality rate** — the share of people diagnosed with a disease who die of it. It moves with treatment and with how well cases are counted, which is why cholera's figure fell from 30 to 60% in nineteenth-century cities to 0 to 10% today.
- **Yersinia pestis** — the bacterium that causes bubonic and pneumonic plague. Naming it matters here because fourteenth-century people had no idea it existed, which is Dattani's point about why the Black Death spread unchecked.
- **Antigenic shift** — the sudden recombination of different flu strains into a new one. It is the mechanism that makes influenza pandemics recur rather than happen once.
- **Public Health Emergency of International Concern (PHEIC)** — the WHO's highest formal outbreak declaration, carrying a legal obligation on countries to respond promptly. It exists because "pandemic" itself has no official definition.

Numbers in this reading

- Black Death, 1346 to 1353: around 50 to 60% of Europe's population, approximately 50 million people, in six years.
- Competing Black Death estimates in the endnotes: upwards of 60% (Benedictow), 51 to 58% (Aberth), 40% (Christakos et al.), 39% of Western Europe (Jedwab, Johnson and Koyama).
- Y. pestis emerged at least 4,000 years ago; first plague pandemic began 541 CE, recurring until the mid-eighth century; Justinianic Plague 541 to 549 CE; third pandemic 1894 to 1940.
- Black Death outbreaks continued until around 1690.
- Columbian Exchange: an estimated 48 million deaths; population of about 54 million before 1492, falling to 5.6 million in 1600, a reduction of about 90%.
- 1918 influenza: 50 to 100 million deaths. Endnote alternatives: 30 million globally (Patterson and Pyle, 1991), 18 million in India alone (Mills, 1986), around 17 million (Spreeuwenberg et al., 2018).
- Other flu pandemics: 1889 around 4 million; 1957 around 2 million; 1968 around 2 million; 2009 between 100,000 and 1.9 million.
- Cholera: seven pandemics in two centuries; at least 23 million deaths in India between 1865 and 1947; over 900,000 recorded deaths globally since 1960 in the seventh pandemic; case-fatality 30 to 60%

in major nineteenth-century cities, 0 to 10% globally today.

- HIV/AIDS: 100% fatality rate when identified; median survival about one year after diagnosis; an estimated 33 million deaths between 1981 and 2022; around 1.5 million deaths averted each year by antiretroviral therapy; around 600,000 deaths annually in recent years.
- COVID-19: around 27 million excess deaths from January 2020 to November 2023. To end-2021: 18.2 million (IHME), 14.8 million (WHO), 17.8 million (The Economist).

What the author is unsure about

Dattani is unusually open about her uncertainty, and most of it lives in the endnotes. She says there is no universally accepted definition of a pandemic. She says comprehensive global death tolls are simply not available for the Black Death, for cholera, or for many events on the timeline, which is why some entries appear as triangles. She names the specific weaknesses of the Black Death method: the net-change approach can absorb deaths from other events of the period, and the underlying records come from only parts of four or five European countries. She flags that pre-Columbian population estimates rest on important assumptions, such as how many houses in a settlement were occupied at a given time. She says the 1918 range of 50 to 100 million reflects missing

death registration, missing records, misdiagnosis and under-reporting in rural and native populations, and that a recent lower estimate of around 17 million is likely itself an underestimate. She notes that cholera case-fatality figures are limited by weak surveillance and by countries using different case definitions. And she shows wide uncertainty around the COVID-19 excess-death estimate while arguing that even its floor is far above the confirmed count. Her choice of The Economist over the WHO and IHME is explained as a judgement about updating and documentation, not as a finding.

For the discussion

- A point worth making: this reading is really about measurement, and the choice of method changes the headline. COVID-19 looks like 7 million deaths through confirmed counts and around 27 million through excess mortality, and only one of those two numbers can be compared with the 1918 figure.
- A tension worth raising: every historical toll in the piece has been revised upward as records were digitised, and Dattani explains why (earlier records over-represented the nobility, who were better fed and healthier). If our past keeps getting worse as we look harder, what does that do to the base rates the rest of the course will use?
- A question worth asking: the pandemics with no death toll at all are drawn as triangles rather than

circles. How should a course about worst-case planning treat the events we cannot even size?

1.2-2 · FURTHER · COURSE TIME 20 MIN

An illustrated history of the world's deadliest epidemics, from ancient Rome to Covid-19

Erik English · 2023

[Read the original paper](#) → · the course assigns: the whole piece, from a structured account another helper made while reading it, not the original text

A Bulletin of the Atomic Scientists photo-essay walks through every known epidemic that killed more than a million people, and argues that epidemics are not random weather but a consequence of how we live.

Average summary

Summarized from a structured account rather than the original text, so nothing here is a direct quotation: raw/1.2-2.txt is a dense record another helper wrote while reading the feature in a browser, keeping its structure and its full table of 19 epidemics with dates, pathogens and death tolls, which are used exactly as given.

The piece sets itself one rule: it covers known historic epidemics and pandemics that resulted in more than

one million deaths, and leaves some out where the data is too uncertain or unavailable. It groups them under Antiquity, Pre-industrial and Modern.

Antiquity: the Antonine Plague from 165 (Roman Empire, likely smallpox), 5 to 10 million dead, a quarter to a third of the Roman population. The Plague of Justinian from 541, an estimated 100 million dead by 590, with Constantinople losing roughly a fifth of its citizens at 10,000 deaths a day. The Japanese smallpox epidemic from 735, which killed a third of Japan's population by 737.

Pre-industrial: the Black Death from 1346, more than 25 million dead between 1347 and 1351, and plague more than 200 million worldwide across all of history. The 1520 Mexico smallpox epidemic, five to eight million. Two cocoliztli epidemics, 1545 to 1548 and 1576; by 1600 Mexico's population had fallen to only two million from 15 to 30 million at contact. The Italian plague of 1629 to 1631, roughly one million in northern Italy. The Naples plague of 1656, an estimated 1,250,000 across the Kingdom of Naples. The Persian plague of 1772 to 1773, an estimated two million.

Modern: the third cholera pandemic (1846 to 1860), with more than a million believed dead in Russia alone and no global estimate available. The third plague pandemic from 1855, more than 15 million deaths, 12 million of them in India under the British

Raj, ending in 1960. The 1889 to 1890 flu, an estimated one million, with St Petersburg losing 180,000 of its one million people within five weeks. The 1918 influenza pandemic, 40 to 50 million deaths from 1918 to 1920, more than all deaths in the First World War. The Russian typhus epidemic of 1918 to 1922, two to three million deaths out of 20 to 30 million cases. The 1957 flu (one to four million), the 1968 flu (one million). HIV/AIDS, more than 40 million dead as of 2021. COVID-19, 6.7 to 28.1 million.

The closing argument is the piece's real thesis. It puts the annual risk of another COVID-scale pandemic at about two percent and says that estimate is expected to triple, because human activity is raising the risk of zoonotic spillover. Quoting David Quammen's *Spillover*, it argues outbreaks reflect things we are doing, not only things happening to us: climate change, conflict and environmental degradation all contribute. Technically we are more capable than ever (COVID-19 vaccines were approved in record time and saved upwards of 20 million lives in 2021 alone), yet the circumstances that produced COVID-19 remain largely unchanged.

Why it's assigned

It is the Further-tier companion to 1.2 #1, and it does something the Our World in Data piece does not: it tells each epidemic as a story, with the politics, trade routes and war that carried the disease. For the day's

first question (how bad have pandemics been?) it supplies a second, independent set of tolls to compare against, and for the day's second question it supplies the mechanism: empires, trade, industrialisation, war, and now air travel and ecological damage.

Words to know

- **Virgin soil epidemic** — an outbreak in a population with no prior exposure and so no immunity. The piece uses it for Japan in 735 and Mexico in 1520, and it explains why those two tolls are so extreme.
- **Cocoliztli** — the Nahuatl word for pestilence, used for the 1545 and 1576 epidemics in New Spain whose pathogen is still unidentified; leading theories are salmonella or a viral haemorrhagic fever tied to drought and living conditions under Spanish rule.
- **Quarantine** — from the Italian *quaranta giorni*, forty days: the waiting period Venice required of ships arriving from infected ports during the Black Death. The word is a fossil of the first containment policy.
- **Excess deaths** — deaths above the expected baseline. The piece uses it for the gap between COVID-19's official 6.7 million and an estimate closer to 21 million.

- **Zoonotic spillover** — a pathogen crossing from an animal into humans. It is the mechanism behind the closing argument that our own activity sets the risk.

Numbers in this reading

- Inclusion rule: epidemics with more than one million deaths.
- Antonine Plague, from 165: 5 to 10 million dead, between a quarter and a third of the Roman population; upwards of 2,000 deaths per day in Rome by some accounts.
- Plague of Justinian, from 541: an estimated 100 million dead by 590; 10,000 deaths a day in Constantinople, roughly one fifth of its citizens.
- Japanese smallpox, 735: a third of Japan's population dead by 737.
- Black Death, from 1346: more than 25 million dead between 1347 and 1351; plague has killed more than 200 million worldwide across history.
- Mexico smallpox, 1520: five to eight million. Cocoliztli 1576: a further two to two-and-a-half million; Mexico's population down to two million by 1600 from 15 to 30 million at contact.
- Italian plague 1629 to 1631: roughly one million in northern Italy. Naples plague 1656: an estimated 1,250,000 across the kingdom, 150,000 to 200,000 in the city, more than half the city's population.

- Persian plague 1772 to 1773: an estimated two million.
- Third cholera pandemic, 1846 to 1860 (originating in Bengal in 1839): more than one million in Russia alone, no global estimate.
- Third plague pandemic, from 1855: more than 15 million worldwide, 12 million in India; ended 1960 when annual deaths fell below 200.
- 1889 to 1890 flu: an estimated one million; St Petersburg lost 180,000 of one million inhabitants within five weeks of November 1889.
- 1918 influenza: 40 to 50 million deaths, 1918 to 1920.
- Russian typhus, 1918 to 1922: two to three million deaths out of 20 to 30 million cases.
- 1957 flu: one to four million worldwide; 116,000 American deaths; vaccine given to 30 million Americans. 1968 flu: one million worldwide, 1968 to 1970; an estimated 160 million people travelled by commercial airliner during it.
- HIV/AIDS: more than 40 million dead as of 2021. COVID-19: 6.7 to 28.1 million, with official count 6.7 million and excess-death estimates closer to 21 million.
- Annual risk of another COVID-scale pandemic: about two percent, expected to triple. COVID-19 vaccines saved upwards of 20 million lives in 2021 alone.

What the author is unsure about

English says plainly that pinning down exact tolls is hard even in an analytics-heavy age, and that wide uncertainty ranges are the historical norm; the only certainty is that more pandemics are coming. Specific admissions: no global death-toll estimate exists for the third cholera pandemic; the Japanese 735 toll is derived from records of loan defaults on rice purchases, assumed to indicate the borrower's death, and is treated as conservative; the Italian plague estimates rest partly on Alessandro Manzoni's novel *The Betrothed*; some scholars consider the Justinian estimates and their consequences exaggerated; recent theories argue human body lice, not rat fleas, were the main Black Death vector; the cocoliztli pathogen remains unidentified; and whether SARS-CoV-2's spillover was natural or the result of a laboratory accident is left as unclear. The two percent annual risk and the expectation that it will triple are forecasts, not findings.

For the discussion

- A point worth making: the piece's own reader comments name its biggest gap, which is that China, India, Africa and indigenous deaths in the Americas are thinly covered, and one commenter notes tuberculosis is not mentioned at all. A list of the deadliest epidemics is also a list of whose records survived.

- A tension worth raising: this reading and 1.2 #1 give different tolls for the same events, for example 40 to 50 million versus 50 to 100 million for 1918, and more than 25 million versus around 50 million for the Black Death. Which number a policy argument reaches for is itself a choice.
- A question worth asking: the piece links outbreaks to xenophobia and discrimination, and shows typhus fear in Eastern Europe being channelled into anti-Semitism and used as Nazi propaganda. If social response is part of the harm, should it be part of the defence plan the course is building?

1.2-3 · FURTHER · COURSE TIME 120 MIN

A Short History of Biological Warfare: From Pre-History to the 21st Century

W. Seth Carus · 2017

[Read the original paper →](#) · the course assigns: from the introduction, the three period summations, the closing summary and "Biological Warfare's Future", plus every section heading and selected pages on Japan and the Soviet programme

Carus's survey of biological warfare from prehistory to the present reaches a deflating conclusion: resort to biological weapons has been rare, the only use causing substantial loss of life was Japan's campaign against

China in the 1940s, and nobody can confidently say why.

Average summary

From the introduction, the three period summations, the closing summary and the final section "Biological Warfare's Future", plus the chapter headings and the pages on the Japanese campaign and the Soviet programme. Carus writes for the National Defense University's Center for the Study of Weapons of Mass Destruction.

He splits the history into three periods. Before 1900, that is before science proved microorganisms cause disease, resort to biological warfare was **exceedingly rare**, and he tells readers interested only in modern history they can skip that section. Many famous episodes turn out to be misreported: a 1784 to 1785 incident in Tunis, where plague victims were buried near the European enclave and pus-soaked cloth thrown into it, is wrongly retold elsewhere as a besieging army catapulting contaminated clothing. A Confederate physician's plan to spread yellow fever with contaminated clothing could not have worked, because yellow fever is mosquito-borne, as was only shown in 1900.

From 1900 to 1945 came state programmes, use in both world wars, and the 1925 Geneva Protocol. The first organised state campaign was German sabotage

in the First World War, aimed at animals. The most extensive use was Japan's attacks in China: almost all known victims of biological warfare were Chinese, mostly civilians. In the 1942 Zhe-Gan campaign the Japanese spread *Yersinia pestis* by aircraft, contaminated water supplies with cholera, typhoid and paratyphoid organisms, spread *Bacillus anthracis* in rice fields, released plague-infected fleas and handed out contaminated cakes. Estimates of the overall toll range from one Chinese study's 740,000 to a Western expert's "a few tens of thousands"; either way these were history's most lethal uses of biological weapons.

After 1945, superpower investment turned a sabotage technique into one capable of mass casualties, largely through aerosol dissemination. Carus reports a claim that the Soviet Union held 20 tons of smallpox virus and built three plants with an annual wartime capacity of 1,800 tons of *B. anthracis*. But lethality did not translate into use: there is no evidence of widespread use since 1945, only small-scale sabotage, none of it exploiting the new technology. The deadliest biological attacks since the war, he notes, have been attributed to criminals.

His closing question is why. He sets out two camps. Some argue that advances in biology, globalisation of skills, and growing access to enabling technology will inevitably bring more and deadlier use, possibly putting Cold War superpower capability in the hands

of lone actors. Others argue biological weapons are harder to develop and employ than claimed, that tacit knowledge (undocumented know-how held by a shrinking group of practitioners) is the real constraint, and that strong norms make use morally or politically repugnant. Carus's own verdict: if use increases, it will be because some past constraint has disappeared, and the most likely cause is not technology but a fundamental change in attitudes toward using disease as a weapon.

Why it's assigned

It is the long Further-tier history behind the day's first question, and it is the most sceptical voice on the 1.2 page. Where the other readings emphasise rising capability, Carus's record shows capability and use coming apart for a century. That makes it the natural counterweight to the 1.3 readings on AI and synthetic biology: before the course argues the ceiling is rising, this reading asks why the historical floor has stayed so low.

Words to know

- **BW** — Carus's abbreviation throughout for biological warfare. Worth knowing because the paper uses it in almost every sentence.
- **Biological agent versus toxin** — an agent is a replicating biological entity such as a bacterium; a toxin is a poison of biological origin, closer to a

chemical weapon, and banned by the Chemical Weapons Convention as well. The monograph focuses on what the Biological Weapons Convention covers.

- **Tacit knowledge** — undocumented know-how needed to actually make something work, held by practitioners rather than written down. Carus presents it as possibly the real barrier to building biological weapons, more than published science.
- **1925 Geneva Protocol** — the first arms-control instrument covering biological methods of warfare; it essentially prohibited first use rather than possession.
- **Biological and Toxin Weapons Convention (1972)** — the treaty banning development, production and stockpiling. It is the dividing line Carus uses for his modern period.
- **Aerosol dissemination** — spreading an agent as a breathable cloud. This is the specific technical advance that, in Carus's telling, made biological weapons potentially as deadly as thermonuclear ones.
- **Biological sabotage** — small-scale covert infection, typically of animals or supplies, rather than mass attack. Carus's point is that nearly every real use, before and after 1945, has been at this scale.

Numbers in this reading

- Zhe-Gan campaign, 1942: 175,000 Japanese troops against about 260,000 Chinese troops; the Japanese reportedly infected 3,000 Chinese prisoners of war with typhoid and paratyphoid organisms and freed them to spread disease.
- Japanese losses in that campaign: 1,000 dead and 10,000 wounded or sick from all causes, against a claimed 30,000 Chinese military losses; a senior US officer estimated as many as 250,000 Chinese died as a result of the campaign.
- A Chinese account claims the biological attacks made 1 million people sick, killing a "few" tens of thousands; Carus says this cannot be verified and the real numbers might have been far lower.
- Japanese blowback: 10,000 of their own troops sickened, 1,700 killed, according to one account.
- Overall Japanese biological warfare toll: one study puts it at 740,000; a Western specialist concludes "a few tens of thousands of deaths overall, almost all Chinese civilians".
- Soviet programme: a claimed stockpile of 20 tons of smallpox virus, constantly replenished; three plants with an annual wartime production capacity of 1,800 tons of B. anthracis; 14 production and mobilisation plants and 8 special weapons and facility design units; capacity of hundreds of tons of agent per year.
- US programme spending: about \$1.9 billion, over \$25 billion in later terms.

- 2005 US State Department judgement: "Russia continues to maintain an offensive BW program in violation of the Convention." In 2016 the US reported Russia had "not satisfactorily documented whether this program was completely destroyed or diverted to peaceful purposes".

What the author is unsure about

Carus's uncertainty is the spine of the paper, not a footnote to it. He states at the outset that much about biological warfare remains unknown and may stay unknown, and that this is an overview rather than a definitive history. He does not resolve the Japanese death toll, presenting the 740,000 figure and the "few tens of thousands" figure side by side and saying it is impossible to verify. He says it is unclear why biological warfare has been so rare, given that simple biological sabotage has been accessible for more than a hundred years, and that there is no clear answer to why there has been so little use since 1945, suggesting a combination of lack of interest, countermeasures, and "at least some good luck". He says it is impossible to assess the likely effectiveness of Soviet weapons because so little is known about their design, concepts of operation or intended strategic role; and unclear whether Russia has entirely terminated the programme. He gives the optimistic and pessimistic cases about the future as competing

expert positions, not as his finding, and ends on a conditional rather than a prediction.

For the discussion

- A point worth making: the single most lethal use of biological weapons in history was carried out with crude means by a state with, in Carus's words, inadequate scientific and engineering foundations. Capability and casualties have never tracked each other neatly.
- A tension worth raising: Carus argues the binding constraint may be attitudes rather than technology. If that is right, then DNA synthesis screening and model safeguards are working on the wrong variable, or at least not the decisive one.
- A question worth asking: he reports that the deadliest biological attacks since 1945 have been attributed to criminals rather than states or terrorists. Does the course's threat model have a place for that?

1.2-4 · CORE · COURSE TIME 20 MIN

Germ Warfare: A Very Graphic History

US Bipartisan Commission on Biodefense (written by Max Brooks) · 2019

[Read the original paper →](#) · the course assigns: whole book, from the publisher's audio edition

Max Brooks tells the whole history of biological warfare, from Scythian archers to the 2001 anthrax letters, in order to make one argument: the thing that keeps a population healthy is the same thing that keeps it safe, and we are letting it decay.

Average summary

Summarized from the publisher's audio edition of the graphic novel, because the PDF is artwork. A few proper names come through the narration imperfectly, and where the standard spelling is well known it is used here. Written by Max Brooks, narrated by James Lewis, with a guest performance by Erica Lewis, executive producer the Bipartisan Commission on Biodefense, copyright 2019.

The frame is simple. Microbes are everywhere, and the dangerous ones have **killed or crippled more human beings than all the wars in history**. For most of human history we had no idea where disease came from, so cures ran from useless to murderous, and the mass slaughter of cats in pre-modern Europe may have raised the rat population and so helped spread plague-infected fleas. People did not know what caused disease, but they knew how deadly it was, and that is why they tried to use it.

Then the episodes, in order: Scythian archers dipping arrows in blood and horse dung; Mongol warriors

catapulting plague-infected corpses over the walls of Genoese Kaffa in 1346; the British commander at Fort Pitt handing a Delaware peace delegation blankets and a handkerchief from the smallpox hospital in 1763; a Confederate sympathizer stealing yellow fever victims' clothing in Bermuda in 1864 to sell at auction in the United States. The nineteenth century brings Semmelweis, Lister, Pasteur, Elizabeth Blackwell, Dorothea Dix, Clara Barton, Mary Walker and the army of nurses of the American Civil War. Then the line that organises the book: "while science helped win the war on germs, it also helped use germs for war."

Imperial Germany ran a global biological sabotage campaign in the First World War, mostly against the horses and mules that moved armies, including from a plant in a basement in Silver Spring, Maryland. That scare helped put germs into the 1925 Geneva Protocol. Japan did not ratify it, and General Shiro Ishii, reasoning that a weapon worth banning must be worth having, built the largest and most lethal germ warfare campaign the world has ever seen. Two facts emerged, Brooks says: Chinese areas with stronger public health systems withstood the attacks better, and the world finally understood that microbes could be weapons of mass destruction.

The Cold War ran a secret biological arms race alongside the nuclear one, with viruses the most attractive agents because antibiotics do not stop

them. Here Brooks makes his central claim: unlike nuclear war, defence against germs did not force a choice between spending on peace and spending on war, because what kept people healthy also kept them safe. Nixon renounced the US offensive programme in 1969, the Biological Weapons Convention followed in 1972, and the Soviet Union violated it, as the Sverdlovsk anthrax release showed.

The book ends on non-state actors, the two successful bioterrorist attacks on American soil, and then on us: public health has become a victim of its own success, and conspiracy theories, budget cuts and apathy threaten to erode everything we built.

Why it's assigned

The course's own note says this graphic novel chronicles biological warfare from ancient Greek archers to Cold War weapons programmes, showing how nation states and terrorist groups have repeatedly weaponised disease. It is the fastest way to get the whole history in one piece, and it hands Unit 1 two ideas the rest of the week keeps returning to: that public health infrastructure is itself a defence against deliberate attack, and that the barrier to misuse is not raw materials, since microbes are everywhere and can be brewed cheaply, unlike uranium.

In depth

Before germ theory

The book opens on the invisible: microbes, bacteria, viruses, most harmless or helpful, the dangerous ones deadlier than all of history's wars. Because people had no idea microbes existed, treatments ranged from useless to absurdly harmful (**bleeding, leeches, drilling holes in the head**) to hysterically murderous, and panic found scapegoats: the narration dramatises crowds blaming Roma and Jewish people for the plague and calling for them to be killed. Antonie van Leeuwenhoek's discovery of "animalcules", an early term for microscopic life, is the turning point in understanding. Armies already knew enough to fear infection, which killed more soldiers than the enemy did.

Early attempts to use disease

Scythian archers dipped their arrows in a mixture of blood and horse dung to make sure everyone they shot became infected.

In 1346, Mongol warriors catapulted plague-infected corpses over the walls of the Genoese city of Kaffa.

In 1763, during the siege of Fort Pitt, the British garrison commander William Trent gave a peace delegation of Delaware Native Americans two blankets and a handkerchief out of the smallpox hospital, to convey smallpox to the Indians. The book cites the journal of William Trent and an invoice to the

crown of August 1763. It says plainly that it is unknown how effective this biological attack was, because the Delaware were already suffering the ravages of European smallpox.

In 1864, during the American Civil War, the Confederate sympathizer Dr Luke Pryor Blackburn volunteered at a yellow fever treatment hospital in Bermuda so he could steal the infected clothing, smuggle it back to the United States and sell it at auction, intending to destroy the army and everyone in the country. The source given is the testimony of Blackburn's agent, Godfrey Joseph Hyams. The trunks were never opened and the public was never exposed.

Public health arrives, and so does germ warfare

By the nineteenth century the medical and public health communities began to get their act together: Ignaz Semmelweis; Sir Joseph Lister, with carbolic acid disinfecting both the wound and the surgical instruments; Louis Pasteur, with the vaccine that helps the body build resistance; and women including Dr Elizabeth Blackwell, Dorothea Dix, Clara Barton, Dr Mary Walker and the army of nurses who served in the American Civil War. The sanitation lesson is played as a joke: you cannot dig your toilets right next to the creek you drink from.

Then the hinge: science helped win the war on germs and also helped use germs for war. The First World

War (1914 to 1918) saw submarines, airplanes, machine guns, poison gas and germs. Many countries experimented, including Imperial Germany, and unlike earlier guesswork this was based on hard science, growing weaponized bacteria including *Pseudomonas mallei* and anthrax. Horses were the primary target because most of the world's armies depended on them for transportation. Germany ran a global campaign of biological sabotage, attempting to infect animal populations in Finland (then still part of Russia), France, Argentina and the USA, and German saboteurs set up their own biological weapons plant in the basement of a home in Silver Spring, Maryland. Brooks says it is hard to know how effective the programme really was, but it frightened world leaders into including germs in the 1925 protocol for the prohibition of the use in war of asphyxiating, poisonous or other gases, and of bacteriological methods of warfare.

Japan

Not all nations ratified the Geneva Protocol, including the rising power of Imperial Japan. Under General Shiro Ishii, Japan conducted the largest, most lethal germ warfare campaign the world has ever seen. The irony Brooks points out is that Ishii was inspired by the Geneva Protocol: he reasoned that if the rest of the world wanted to ban germ weapons, they must have serious potential. The largest plant was in the

Ping Fan district of the occupied puppet state of Manchukuo.

The arsenal: smallpox, anthrax, glanders, tularemia, tuberculosis, tick encephalitis, gas gangrene, brucellosis, tetanus, botulism, salmonella, cholera, typhoid, typhus and bubonic plague. Like their Nazi allies, the Japanese used human beings as live test subjects; most victims were Chinese civilians, but they also experimented on British, Australian and American prisoners of war. The estimated number of live victims ranges between 3,000 and tens of thousands.

The attacks against the Soviet Union and China: pouring infected slurry into rivers, spraying rice paddies, giving out infected cakes, giving prisoners infected food and then releasing them in the hope they would infect their families, directly infecting prisoners and sending them across the border to infect Russians. Aircraft dropped food packages infested with plague-carrying fleas, and porcelain bombs packed with infected fleas or tiny black particles of concentrated plague.

The toll is uncertain: accounts range from tens of thousands to one Chinese study placing the dead at roughly 740,000. Brooks says two facts emerged regardless of the exact number. First, Chinese areas with stronger public health systems were better able to withstand the germ attacks. Second, the world

finally woke up to the fact that microbes could be used in weapons of mass destruction.

The Cold War's second arms race

While the Cold War is largely seen through the nuclear lens, a second biological arms race was waged in secret in the US, the Soviet Union and many other nations. Viruses became the most attractive germ weapon because they are smaller than bacteria, can get into human cells before the immune system detects them, and are immune to antibiotics, the new wonder drug of the early twentieth century. Since antibiotics will not stop a virus, the only sure defence is preventing infection with a vaccine.

This leads to the book's thesis. Vaccines, along with well-funded labs, healthcare, education and all the other elements of a strong public health system, did more to protect against a germ war than anyone could have done about a nuclear war, because with germs you do not have to choose between spending money on peace or on war. In the case of a germ attack, what kept us healthy also kept us safe.

Treaties and violations

On 25 November 1969, in the Roosevelt Room at the White House, President Richard Nixon announced the end of the US offensive biological weapons programme: "The United States shall renounce the

use of lethal biological agents and weapons and all other methods of biological warfare."

In 1972 the US, the Soviet Union and over 100 other countries signed the Biological Weapons Convention, banning the development, stockpiling, acquisition, transfer, retention, assistance with and production of biological agents, weapons, equipment and delivery systems.

In late 1979, rumours leaked out of the USSR about an anthrax epidemic in and around Sverdlovsk. Western intelligence suspected an accident at the Sverdlovsk bioweapons plant; the Soviet government blamed tainted meat. Rumours kept coming, along with smuggled photos of military checkpoints and decontamination trucks around the facility. President Boris Yeltsin finally admitted the truth in a 1992 interview with Komsomolskaya Pravda, dated 27 May 1992. The full details of the release have yet to be fully disclosed, and Russia refuses to let the US into some of its biological research facilities, insisting, as the US does, that its programme is purely defensive.

Brooks then widens the frame: 16 countries have yet to sign or ratify the Biological and Toxin Weapons Convention, and some that have, such as North Korea, are suspected of continuing secret programmes. But he argues the greatest danger may come from non-state actors. A biological weapon is more attractive to them than a nuclear one, because uranium can only

be mined in 20 countries and is incredibly expensive and difficult to process, while microbes exist everywhere on Earth and can be brewed cheaply at home.

The two attacks on American soil

To date, the book says, there have been only two successful bioterrorist attacks on American soil.

The first, in 1984, was carried out by the Rajneeshee movement in Wasco County, Oregon. The cult's bioweapons programme was spearheaded by a nurse the narration names as Ma Anand Puja, and on the orders of the cult's second-in-command, Ma Anand Sheela, cultists infected several restaurants in Wasco County with Salmonella. The goal was to sicken enough people to keep them from voting in a local election. Because none of the victims died, because officials feared publicity would lead to copycat attacks, and because the attacks took place a decade and a half before September 11, the perpetrators were charged as criminals, not terrorists. Most conspirators were given light sentences, while Rajneesh himself, who pleaded innocence to his death, was simply asked to leave the United States.

In 2001, one week after the September 11th attacks, envelopes containing what appeared to be military-grade anthrax were mailed to several locations across the US, with threatening notes. Targets included

members of Congress and news media. Twenty-two people were sickened; of these, five were killed, including two postal workers. The FBI's Amerithrax investigation eventually led to Dr Bruce Edwards Ivins, an American microbiologist working for the US Army at Fort Detrick, who died by suicide in 2008 before he could be charged. The investigation took nearly a decade; the public panic was immediate.

Brooks quotes former White House Press Secretary Scott McClellan's memoir: President Bush's thinking was deeply affected by the anthrax attacks, and he was determined not to let another terrorist attack happen on his watch and to challenge regimes believed to be seeking weapons of mass destruction. Critics charge that the bioterrorist threat was used to manipulate Americans into supporting the Iraq War; defenders claim intelligence assessed that Saddam Hussein had biological weapons. No biological, chemical or nuclear weapons were ever found by US forces or United Nations inspectors in Iraq, and the book says the perception of crying wolf damages the public trust to this day.

What we have, and what we are losing

Globalization has empowered the modern bioterrorist: online information is readily accessible, as is lab equipment that can be ordered from anywhere in the world with one click.

Against that, Brooks lists the ordinary defences already in place, drawn as everyday signage: childhood vaccination, "employees must wash hands", covering your mouth when you cough, USDA meat inspection, restaurant sanitary grades, Grade A pasteurized milk, food inspectors, wastewater treatment, and public health advice during Zika. Just as in the Cold War biological arms race, what keeps us healthy can also keep us safe.

But he says we can and should do more: increased investment in biological detection technology, enhanced communication between public health and national security organizations, leadership and coordination by a single high-level official in the White House, increased congressional oversight, and greater, more effective public outreach. The response is dramatised through a full cast (the White House, Secretary of Defense, National Security Advisor, Secretary of Homeland Security, Secretary of State requesting an emergency UN meeting, Attorney General, Chairman of the Joint Chiefs of Staff, Secretary of Health and Human Services saying the private sector can ramp up vaccine and protective gear production within the hour, Homeland Security Advisor and Vice President) and named facilities (USAMRIID at Fort Detrick, the CIA, the CDC in Atlanta).

The final turn is the one that gives the book its edge. The terrorist's greatest ally is us. Public health has become a victim of its own success: at the turn of the last century infectious diseases were the number one killers of people in the industrialized world, and now they do not even rank in the top five. So most people take public health for granted and some do not see why we need it any more. Brooks shows the distrust coming from the left, the centre and the right alike, and names conspiracy theories, budget cuts and personal apathy, anti-vaxxer clusters and CDC budget cuts, as what threatens to erode all the progress humanity has made. Germs will always be with us, and we must always be ready. If we accept the cost in time and money, we will never have to pay the cost in lives.

The credits note that the Bipartisan Commission on Biodefense was established in 2014 to comprehensively assess US biodefense efforts and issue recommendations to foster change. Its 2015 report, "A National Blueprint for Biodefense: Leadership and Major Reform Needed to Optimize Efforts", identified capability gaps and recommended changes to US policy and law. Former Senator Joe Lieberman and former Governor Tom Ridge co-chair the Commission, joined by former Senate Majority Leader Tom Daschle, former Representative Jim Greenwood, former Homeland Security Advisor Ken Wainstein and former Homeland Security and

Counterterrorism Advisor Lisa Monaco. The Hudson Institute is the Commission's fiscal sponsor.

Words to know

- **Animalcules** — the early word for microscopic life, used by Antonie van Leeuwenhoek. It marks the moment people could see the cause of disease at all.
- **Glanders** — a bacterial disease of horses and mules, one of the agents Imperial Germany grew. It matters here because the First World War campaign targeted transport animals, not people.
- **Biological sabotage** — covert infection of animals, supplies or food rather than mass attack. Nearly every real episode in the book is at this scale.
- **Weaponized** — an agent prepared and formulated so it can actually be delivered and cause harm, as opposed to simply grown. The 2001 letters are described as containing what appeared to be military-grade anthrax.
- **Amerithrax** — the FBI's name for its investigation into the 2001 anthrax letters. Worth knowing because it is how that case is referred to in almost all later reading.
- **Non-state actor** — an individual or group acting outside any government. The book's argument is that these are now the greatest danger, because the inputs for biology are cheap and everywhere.

Numbers in this reading

- 1346: Mongol warriors catapult plague-infected corpses into Kaffa.
- 1763: two blankets and a handkerchief from the smallpox hospital, given at the siege of Fort Pitt; sources dated August 1763.
- 1864: the Bermuda yellow fever clothing plot; the trunks were never opened.
- 1914 to 1918: the First World War, with German biological sabotage in Finland, France, Argentina and the USA.
- 1925: the Geneva Protocol prohibits bacteriological methods of warfare.
- Japan's human test subjects: between 3,000 and tens of thousands.
- Japan's biological warfare death toll: tens of thousands, up to roughly 740,000 in one Chinese study.
- 25 November 1969: Nixon ends the US offensive biological weapons programme.
- 1972: the Biological Weapons Convention signed by the US, the Soviet Union and over 100 other countries.
- 1979: the Sverdlovsk anthrax epidemic; Yeltsin's admission comes in a 27 May 1992 interview.
- 16 countries have yet to sign or ratify the Biological and Toxin Weapons Convention.
- Uranium can be mined in only 20 countries.

- 1984 Rajneeshee attack: several restaurants in Wasco County contaminated with Salmonella; no deaths.
- 2001 anthrax letters: 22 people sickened, 5 killed including two postal workers; Ivins died in 2008; the investigation took nearly a decade.
- Infectious diseases were the number one killers in the industrialized world at the turn of the last century; now they do not rank in the top five.
- The Bipartisan Commission on Biodefense was established in 2014; its national blueprint report came in 2015.

What the author is unsure about

Brooks marks his gaps as he goes. It is unknown how effective the 1763 Fort Pitt attack was, because the Delaware were already suffering European smallpox. It is hard to know how effective Imperial Germany's programme really was. The number of Japanese human test subjects is a range of 3,000 to tens of thousands, and the death toll a range from tens of thousands to roughly 740,000, with no attempt to settle it. The full details of the Sverdlovsk anthrax release have yet to be disclosed, and whether Russia's current programme is defensive cannot be verified from outside. North Korea is "suspected" of continuing a secret programme. On Iraq he presents the critics' and the defenders' accounts side by side and states only the verifiable fact that no weapons

were found. What he does not hedge is the prescription: more detection investment, better communication between public health and national security, one senior White House coordinator, more congressional oversight and better public outreach are offered as the Commission's recommendations rather than as findings.

For the discussion

- A point worth making: the single clearest protective finding in the book is not a weapon or a treaty, it is that Chinese areas with stronger public health systems withstood Japan's germ attacks better. That is the whole case for treating public health as national defence.
- A tension worth raising: the book was made by a biodefense commission, and its closing villains are budget cuts and public distrust. That is honest about where the Commission thinks the risk sits, but it also means the reader is being asked for a specific appropriation.
- A question worth asking: the 1984 Oregon attack was prosecuted as crime rather than terrorism partly because officials feared that publicity would invite copycats. Where should the course land on that trade-off between public warning and giving a playbook away?

Historical Overview: From poisoned darts to pan-hazard preparedness

James W. Martin and George W. Christopher and Edward M. Eitzen · 2007

[Read the original paper →](#) · the course assigns: from the introduction, the summary, and every section heading, plus selected pages on the Soviet and Iraqi programmes, bioterrorism, and the policy sections

The opening chapter of the US Army's *Medical Aspects of Biological Warfare* runs the whole history of biological weapons and then argues that the right response is not a list of named pathogens but a flexible, threat-agnostic capability that works equally well against a natural epidemic.

Average summary

From the introduction, the chapter's own summary and every section heading, plus selected pages on the Soviet and Iraqi programmes, bioterrorism and the policy sections. Two notes on the file: the PDF's byline is George W. Christopher, **Daniel M. Gerstein**, Edward M. Eitzen and James W. Martin, and the text discusses events through 2013, so the saved chapter is a later edition than the registry's 2007 date.

Its sections: Introduction; Early Use; The World Wars; The US Program; The Soviet Program; The Special

Case of Iraq; Other National Programs; Biocrimes; Biological Terrorism; Solutions: Toward Pan-Hazard Preparedness (with subsections on the Biological Weapons Convention, smallpox preparedness, dual use research of concern, and pan-hazard preparedness); and Summary.

The framing is medical and sober. The word "toxin" descends from the ancient Greek for arrow poison, which the authors use to show that weapons and biology have always been linked. They list what confounds this history: secrecy, the difficulty of confirming allegations, the lack of reliable microbiological and epidemiological data, and the use of allegations as propaganda or hoax. They name what is changing as a "democratization in the life sciences", in which biotechnology reaches ever more people, some of whom will have ill intent, plus a second worry: that well-intentioned defensive research may itself produce information that could be misused.

The historical verdict is deflationary. Military use before modern microbiology was limited, possibly because other weapons gave more rapid and predictable results. After microbiology arrived, several nations started offensive programmes, but military deployment has been limited and never decisive in armed conflict. With the exceptions of alleged German sabotage in the First World War, Japanese

field trials in the Second, limited South African and Rhodesian deployments, and small-scale covert operations, there are no well-documented biological attacks by nation-states. The likely deterrents: poor tactical utility given the variables in production, storage and delivery; variable incubation periods and host susceptibilities; medical countermeasures; nuclear deterrence; diplomacy; and political vulnerability. Sverdlovsk, the loss of international goodwill toward the United States after Cold War disclosures, and the political consequences of Iraq's 1996 disclosures all show that the liabilities of a state programme have outweighed the strategic advantages.

The chapter's prescription is the part the course cares about: move from a static list of specific pathogens toward a threat-agnostic, capability-based approach using flexible platform technologies that can be adapted to counter novel, unanticipated pathogens; develop broad-spectrum therapeutics; and build response capacities usable against attacks, natural epidemics or other mass-casualty disasters alike. It closes with Joshua Lederberg's line that there is no technical solution to the problem of biological weapons, and that it needs an ethical, human and moral solution if it is going to happen at all.

Why it's assigned

It is the course's lowest-tier option on the 1.2 page, and it is there as the clinician's and the defence planner's version of the history. Two things make it worth the time for the day's third question, what a pandemic-proof world looks like: it is the clearest statement of why biodefence should be built threat-agnostic rather than pathogen-by-pathogen, which is exactly the logic Units 3, 4 and 5 run on, and it introduces dual-use research of concern, which becomes Unit 6's subject.

Words to know

- **Pan-hazard preparedness** — building one set of capabilities that works against deliberate attacks, laboratory accidents and natural outbreaks at the same time. It is the chapter's title idea and the reason it sits in a biosecurity course.
- **Threat-agnostic (capability-based)** — not choosing in advance which pathogen to prepare for, and instead investing in platforms that can be aimed at whatever appears. The alternative, a static list of named agents, is what the authors say failed.
- **Dual use research of concern (DURC)** — legitimate life-science research whose results could be misused to cause harm. The chapter's worked example is the chemical synthesis of poliovirus.
- **Select agent** — a pathogen or toxin on the US regulated list, with Tier I covering those posing the

greatest risk. The chapter tracks how the list and its security rules were revised.

- **Biocrime versus bioterrorism** — a biocrime targets individuals for personal motives; bioterrorism uses biological agents by a non-governmental individual or group to achieve a political or ideological objective. The chapter keeps them separate because the actors, scale and law differ.
- **Cooperative Biological Engagement Program** — the US programme (formerly Nunn-Lugar Biological Threat Reduction) that helped dismantle the former Soviet infrastructure and convert it to peaceful purposes.
- **International Health Regulations (IHR)** — the WHO framework obliging countries to detect and report outbreaks, which the chapter treats as part of biological defence rather than separate from it.

Numbers in this reading

- 184 BCE: Hannibal has earthen pots filled with snakes hurled onto enemy decks against King Eumenes of Pergamum; the victory is attributed to panic rather than envenomation.
- 1346: cadavers hurled over the walls of Caffa, after war broke out in 1343 between the Genoese and the Mongols.
- Sverdlovsk, early April 1979: retired Soviet general Andrey Mironyuk disclosed that safety filters had

not been activated, releasing aerosolized B. anthracis.

- 1992: Yeltsin admits the offensive programme; September 1992, Russia agrees with the US and UK to terminate the inherited programme and allow onsite inspections.
- Cooperative threat reduction results: three large production facilities and dozens of institutes dismantled or converted, and 150 tons of B. anthracis weapons agent destroyed on Vozrozhdeniya Island.
- Iraq, 1990 to 1991: approximately 150,000 US troops received the licensed anthrax vaccine and 8,000 received botulinum toxoid as an investigational new drug. Iraq's 1995 disclosure covered six major sites and production of 19,000 L botulinum (10,000 L weaponized), 8,500 L B. anthracis (6,500 L weaponized) and 2,200 L aflatoxin (1,580 L weaponized). The weapons were deployed but not used.
- Bioterrorist incidents increased markedly after 1985, with peaks in 1998 and 2001; at least 98% of incidents during 2000 to 2002 were hoaxes.
- 1984 Rajneeshee attack: two Wasco County commissioners given contaminated water on 29 August 1984; slurries of Salmonella typhimurium poured into salad bars, salad dressing and coffee creamers at 10 restaurants; 751 cases of enteritis and at least 45 hospitalisations.

- 2001 anthrax mailings: 22 cases of infection, five deaths, approximately 10,000 individuals offered postexposure prophylaxis.
- Aum Shinrikyo: approximately 10,000 members and \$300 million in assets; nine attempted aerosol attacks between 1990 and 1995, three with *B. anthracis* and six with botulinum toxin; the 20 March 1995 sarin attack killed 12 and sent thousands for emergency care.
- The Biological Weapons Convention has 170 member nations, against 189 for the Nuclear Non-Proliferation Treaty and 188 for the Chemical Weapons Convention. Only one allegation has ever been formally registered under the BWC: Cuba against the United States in June 1997 over the crop pest *Thrips palmi*, unsubstantiated by December 1997.
- Smallpox preparedness: more than 400,000 selected military personnel and 38,000 civilian emergency responders and healthcare workers vaccinated; the Strategic National Stockpile holds enough cell-culture-derived vaccine for the entire US population.
- 24 March 2013: a vial of Guanarito virus, a Tier I select agent, reported missing from the University of Texas Medical Branch at Galveston. The select agent Final Rule took effect 5 October 2012.

What the author is unsure about

The chapter is built around what cannot be known. It states at the start that secrecy, unconfirmable allegations, missing microbiological and epidemiological data, and propaganda all confound the history. It says information about state programmes is obscured by secrecy, propaganda and the lack of rigorous data to confirm allegations of use. It reports that Saddam Hussein's reasons for not using the deployed Iraqi weapons are unclear. It says uncertainties exist about activities in Russia, Iran, North Korea and Syria according to a 2013 US State Department report, and uses that together with the miscalculations of the 2002 National Intelligence Estimate to make a wider point: assessing biological weapons programmes is difficult even for highly skilled professionals. On Aum Shinrikyo's failure it gives several competing explanations rather than one. On research risk it is explicit that even with effective risk management, risk never reaches zero and can only be decreased to an "irreducible minimum". And it concedes that technical solutions, while essential, are not sufficient, which is why it ends on ethics.

For the discussion

- A point worth making: this chapter and the day's other histories agree on the surprising part. The documented record of state biological attack is thin, and the authors think that is because the

liabilities of running a programme have outweighed the advantages, not because the science was hard.

- A tension worth raising: the chapter argues for a threat-agnostic capability and in the same breath describes a smallpox preparedness programme built around one named pathogen, with vaccine for the entire US population. Preparedness policy keeps reverting to lists.
- A question worth asking: it uses the synthesis of poliovirus as its worked example of dual-use harm, and notes that the same platform could have been presented as a way to make live attenuated vaccines. If framing and stated intent change how risky a result is in practice, who should be deciding the framing?

1.2-6 • FURTHER • COURSE TIME 5 MIN

High-risk human-caused pathogen exposure events from 1975-2016

David Manheim and Gregory Lewis • 2022

[Read the original paper →](#) • the course assigns: whole paper (version 2), including the event table and the discussion of each event class

Manheim and Lewis compile every publicly known high-risk pathogen exposure caused by human decisions between 1975 and 2016, find 71 of them, and argue the real figure is far

higher because almost every event on the list became known only by accident.

Average summary

A Data Note in F1000Research, version 2, published 8 July 2022. The dataset covers **71 incidents** of accidental or purposeful exposure to, or infection by, a highly infectious pathogenic agent, every one the result of human decisions and mistakes rather than the clinical treatment of a natural outbreak.

The inclusion rules matter as much as the list.

Weapons events start after the 1975 entry into force of the Biological Weapons Convention; laboratory accidents use the same window, partly for consistency and partly because the modern era of lab safety began after the first Asilomar conference in 1973. The agent has to be usable as a bioweapon, or require similar safeguards, which excludes laboratory HIV infections from mishandled specimens, the frequent laboratory-acquired brucellosis cases, infections of healthcare workers treating natural outbreaks, and brucellosis from live vaccines.

The authors sort events three ways: use in an attack; accidents inside a weapons programme; and accidents in research. Laboratory-acquired infections are the most common publicly known type, with moderate confidence, while they think it far less likely that these are the most worrisome.

The headline argument is about under-reporting, made with evidence rather than assertion. Of the 71 events, a 2019 review of PubMed papers found only three. Many of the events with real detail only exist publicly because the now-defunct Sunshine Project filed records requests at a large number of universities between 2003 and 2006, and only a fraction of those requests were ever filled. An anonymous survey in Belgium covering 2007 to 2012 extrapolated roughly 1,000 bio-accident "possible events" a year; none of the three previously publicly reported Belgian laboratory-acquired infections involved Class-3 pathogens, but 19 to 26 of the 74 to 95 additional infections the survey uncovered did. So the authors call their list "at best a generously underestimated lower bound" on the historical rate in the United States, and probably a far greater underestimate elsewhere.

On intentional use, they note it is rare since the BWC and give the sub-categories: alleged military use by Rhodesia in the late 1970s (anthrax during the insurgency in 1978 to 1979, cholera released at least twice into rivers) and alleged use of cholera by the South African military; alleged human testing in the Iraqi and North Korean programmes; and misuse by rogue and non-state actors, including the failed Aum Shinrikyo anthrax attack and the 2001 US anthrax mailings. Their sharpest observation: the only recorded deaths from bioterrorism involved

pathogens from a state bioweapons programme, cultured by a researcher in that programme.

They close with the influenza comparison. None of the events involving the highest-risk pathogens created uncontrolled spread, although one bio-warfare research accident killed more than a hundred people. Influenza is the exception: the 1977 to 1978 release of a 1950s H1N1 strain spread widely, limited only because much of the population over 20 had latent immunity. A near-repeat came in 2004 to 2005, when a company's proficiency-test kits sent 1957 Asian influenza instead of a harmless strain to thousands of laboratories worldwide.

Why it's assigned

It is the only reading on the 1.2 page that gives a base rate rather than a narrative, and it supplies the third source of risk alongside nature and deliberate attack: us, by mistake, in laboratories built to protect us. For the day's second question, how the threat is changing, it is the piece that lets her say something precise about accident risk, and the methodological point (everything we know, we know by accident) carries straight into Unit 2's chokepoint logic.

Words to know

- **LAI (laboratory-acquired infection)** — an accidental exposure and infection during research. The most common publicly known event class in

the paper, and the one with the best, though still poor, data.

- **GCBR (Global Catastrophic Biological Risk)** — a biological event severe enough to threaten civilisation. The paper's whole purpose is to build a lower bound on the human-caused share of that risk.
- **Seroconversion** — the later detection of antibodies showing that an infection happened, often with no symptoms at the time. Many events are known only this way, which is the paper's evidence that exposures occur without any identified cause.
- **Federal Select Agent Program (FSAP)** — the US programme that regulates possession and transfer of dangerous pathogens and reports incidents to Congress. Several of the paper's entries exist only because those reports were obtained by a Freedom of Information Act request.
- **Possible release event** — the FSAP term for an incident where exposure may have occurred. Hundreds are reported a year, and only about 1% lead to infections.
- **Biodefense, bioweapon and research programme types** — the three programme classes the table uses, because the same accident means something different in a weapons plant, a defensive lab and an academic lab.

- **Gain-of-function** — research that gives a pathogen new properties, including greater transmissibility. The paper's final section argues the risks of this work on easily spread pathogens may outweigh the benefits.

Numbers in this reading

- 71 incidents in the dataset; time period 1975 to 2016; version 2 published 8 July 2022.
- Su et al.'s 2019 PubMed review found only three of those 71 events.
- Harding and Byers' surveys of 1979 to 2004 found 155 Hantavirus cases and one death, 143 Brucella cases, 5 Chikungunya (2 asymptomatic), 4 Dengue and 6 SARS-CoV; their 2017 extension to 2015 added one more death, 108 Hantavirus infections, 235 Brucella and 3 Dengue cases, and no new Chikungunya or SARS-CoV.
- An earlier review found most reported bacterial infections in clinical laboratories (84 versus 471) and most viral infections in research facilities (418 versus 181); the later review found 67% of viral LAIs in research facilities and under 10% of bacterial infections.
- Richardson's unpublished summary: 109 LAIs at the CDC from 1947 to 1973.
- Rusnak et al. found 423 infections in US bioweapons programme records at USAMRIID from 1943 to 1969.

- Fort Detrick deaths: William Boyles (1951), Joel Willard (1959) and Albert Nickel (1964), in separate incidents.
- Sverdlovsk, 1979: the table records 100+ dead.
- 1971 Aral smallpox incident (excluded as pre-BWC): stopped by quarantining thousands for several weeks, incinerating several properties, halting all traffic in and out of a city and vaccinating over 50,000 people.
- Rhodesia, late 1970s: thousands infected, 82+ died, per the table.
- 1984 Rajneeshee attack: 751 infections and 45 hospitalisations, no fatalities (excluded from the main count as the agent lacks biowarfare potential by the paper's criteria).
- 2001 anthrax mailings: 22 infected, 5 died.
- 2004 to 2005 influenza proficiency-test error: kits sent to 2,750 US labs and 3,747 international labs; no infections.
- 1999 Oblivskaya, Russia: 69 hospitalised, 6 dead in an alleged accidental release of a synthesized CCHF-like virus.
- 2004 National Institute of Virology, Beijing: 2 researchers infected with SARS-CoV in a BSL-2 laboratory, spreading to 6 or 11 in the community; one researcher's mother, the first identified case, died.

- Belgian anonymous survey, 2007 to 2012: roughly 1,000 bio-accident "possible events" a year extrapolated; 74 to 95 additional LAIs reported, of which 19 to 26 involved Class-3 pathogens; only a quarter of respondents worked in BSL-3 laboratories and Belgium has no BSL-4 laboratories.
- FSAP reports: 1 to 2 infection events per year; hundreds of "possible release" events per year since 2008; on the order of one to two hundred events each year where pathogens are lost in transit, misplaced or cannot be accounted for; only about 1% of reported possible events lead to infections.
- Klotz's estimate, cited: the release into the community of an avian-flu-like pathogen could seed a pandemic with a probability of perhaps 15 percent.

What the author is unsure about

Almost everything, stated plainly. The authors say it is near-certain that there are biological weapons events that are and will remain unknown, because development itself violates the BWC and any deliberate release would be a war crime. They list the incentives against reporting at individual, laboratory and national levels. They say their list is at best a generously underestimated lower bound for the United States and plausibly a far greater

underestimate elsewhere, that it will need revision as new revelations emerge, and that some events will never be known. They have only moderate confidence that laboratory-acquired infections are the most common event type, and think it far less likely that they are the most worrisome. They flag that for the majority of LAIs there appears to be no direct cause, which is itself the evidence that standard practice does not eliminate the risk. They say the relative scarcity of pre-2003 events is more likely a records problem than a real pattern, but also that the later period may not be representative of the earlier accident rate because of the post-2001 expansion of this research. They take no position on the alleged war crimes and human testing. And they note explicitly that better reporting of non-exposure accidents means a rising count of reported possible events may coincide with a steady or falling number of real ones.

For the discussion

- A point worth making: the paper's most quotable finding is not a number but an attribution. The only recorded deaths from bioterrorism involved pathogens taken from a state bioweapons programme by a researcher inside it, which means the defensive enterprise itself has been the deadliest single source.

- A tension worth raising: the authors raise, and call possibly naive, the idea of simply not training and employing scientists to cultivate weaponized strains. A course about building defences should be able to say why that is wrong, or admit that it is not.
- A question worth asking: the 2011 Australian dengue infection happened through a mosquito bite despite correct protective equipment and no researcher error. What does the course's safety model do with accidents that have no cause to fix?

1.3-1 • FURTHER

AI Is Reviving Fears Around Bioterrorism. What's the Real Risk?

Kyle Hiebert · 2025

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

Hiebert argues that chatbots have made bioterrorism newly plausible for non-state actors at exactly the moment the United States is dismantling the institutions that would catch it, and that great-power competition is distracting everyone from the real threat.

[Average summary](#)

The piece opens with a scene. In early 2023 a scientist and former UN weapons inspector, Rocco Casagrande, brought a small container to a White House briefing with US officials. It held a dozen easily available chemical ingredients that Anthropic's chatbot, Claude, had recommended as precursors to trigger another pandemic. Hiebert's reading of the stunt: anyone with an internet connection can now conceivably create their own weapon of mass destruction.

He then does something unusual for a short opinion piece, which is to present the evidence against his own alarm. A team assembled by OpenAI found that **GPT-4 gave users only a marginal advantage over regular internet searches** in building bioweapons. But he argues the risk is not static: the International AI Safety Report produced for the 2025 Paris AI Action Summit shows large language models getting far better at tasks related to biological and chemical weapons, responding accurately to queries about acquiring and formulating deadly agents. Assessments by that report's authors suggest certain models' instructions for releasing lethal substances showed an 80 percent improvement in 2024 alone.

The state-versus-non-state balance comes from the Center for a New American Security's August 2024 report: the unwieldiness and imprecision of bioweapons means states remain unlikely to field

large-scale attacks in the near term, while non-state actors (lone wolves, terrorists and apocalyptic groups) have an unnerving track record of attempting them but with limited success, given the intrinsic complexity of building and wielding such delicate capabilities. CNAS researchers warn that this false sense of security is changing, mostly because of genetic science, synthetic biology and the emergence of cloud labs: discreet, automated facilities contracted to run remote experiments for a client.

The second half is about institutional decay. The Trump White House has gutted America's AI Safety Institute, set up late in the Biden administration to identify, measure and mitigate the risks of advanced AI systems; the International Network of AI Safety Institutes continues similar work with much less sway over American firms. The FBI and CIA, both crucial sources of counterterrorism intelligence, have lost expertise through the intimidation and firing of government employees by the Department of Government Efficiency. And allies, reportedly the Five Eyes nations plus Israel and Saudi Arabia, have grown wary of sharing classified information with Washington out of concern it could reach Moscow. Hiebert notes the inverse risk too: the Kremlin, distracted by Ukraine and dismissive of foreign warnings, was caught flat-footed when Islamic State attacked a suburban Moscow concert hall in March 2024.

He closes on technical fragility and politics. Every major chatbot has proven vulnerable to jailbreaks, and AI agents that execute tasks autonomously amplify this. California's Bill SB 1047 was vetoed by Governor Gavin Newsom after a tech industry campaign, and Republican lawmakers inserted a clause into a tax bill that would ban states and localities from regulating AI for a decade. His last line is a dark joke: if a tech-enabled bioterrorist attack happens, AI will also be vital in expediting a cure.

Why it's assigned

It is the Further-tier opener on the 1.3 page and the most journalistic framing of the day's second question: how are AI and synthetic biology changing the threat? It is useful precisely because it holds two things at once, the evidence that current uplift is marginal and the evidence that the trend line is steep. It also adds a dimension the technical readings leave out, which is that the institutions meant to detect and govern this are being weakened at the same time.

Words to know

- **Jailbreak** — a way users get around a model's safety rules. Hiebert's point is that every major chatbot has been jailbroken, so published safeguards are not the same as working safeguards.

- **Cloud lab** — an automated, remotely contracted laboratory that runs experiments on a client's behalf. It matters because hands-on skill has been the main barrier, and cloud labs let someone rent past it.
- **AI agent** — a system that carries out multi-step tasks autonomously over the internet rather than just answering. Hiebert treats this as the thing that multiplies the jailbreak problem.
- **Non-state actor** — a person or group acting outside any government, which in this piece includes lone wolves, terrorists and apocalyptic groups. The whole argument turns on this category becoming more capable while states stay cautious.
- **AI Safety Institute** — a government body set up to identify, measure and mitigate advanced AI risks. Named here because the American one has been gutted.

Numbers in this reading

- Early 2023: a container of a dozen easily available chemical ingredients, recommended by a chatbot, taken to a White House briefing.
- OpenAI's finding: GPT-4 gave users only a marginal advantage over regular internet searches.
- The International AI Safety Report, for the 2025 Paris AI Action Summit: an 80 percent improvement in certain models' instructions for releasing lethal substances during 2024 alone.

- CNAS report, August 2024, on states versus non-state actors.
- Islamic State attack on a suburban Moscow concert hall, March 2024: 139 people killed, hundreds injured.
- The proposed congressional clause would ban state and local AI regulation for the next decade.

What the author is unsure about

Hiebert does not pretend to settle the question his title asks. He reports OpenAI's marginal-uplift finding and then argues around it rather than refuting it, saying only that the risk is not static. His claim that anyone with an internet connection can conceivably create a weapon of mass destruction rests on one demonstration, and he calls it a stunt. The reporting on allied intelligence-sharing is attributed as what officials "suspect" and "reportedly" involves certain countries. The Zizians are described through a quoted characterisation rather than an assessment. And his closing judgement, that the odds are the next pandemic will be much worse, is an opinion offered as such.

For the discussion

- A point worth making: the strongest single number here is a rate of change, not a level. An 80 percent improvement in one year says more about where to

put safeguards than any snapshot of current capability does.

- A tension worth raising: the piece's two halves pull against each other. If states are unlikely to attack because bioweapons are unwieldy and imprecise, the same unwieldiness should also constrain a lone actor with a chatbot, unless cloud labs and automation really are the difference.
- A question worth asking: he names the weakening of the FBI, the CIA and the AI Safety Institute as part of the threat picture. Should institutional capacity be treated as one of the defensive layers the course is about to build?

1.3-2 · CORE · COURSE TIME 10 MIN

Mother Nature is not 'the ultimate bioterrorist'

Chris Bakerlee · 2021

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

A catchphrase coined after 9/11 to stop people ignoring natural pandemics has been turned around and is now used to dismiss human-made biological risk, and Bakerlee says it should be retired.

Average summary

Bakerlee's target is a single sentence: "Mother Nature is the world's worst bioterrorist." He calls it a **long-overused catchphrase** of scientists and public health professionals in urgent need of retirement, and his argument is historical before it is biological.

The phrase was born in the maelstrom of September 11, 2001 and the anthrax attacks that followed. Its original job was to warn against fixating on bioterrorism while neglecting naturally emerging pathogens, and Bakerlee says that warning proved prescient: over the next several years the war on terror fuelled a tenfold boost in US biodefense spending, and the focus on bioterrorism was so strong that when SARS arrived in 2003 the editorial board of *Nature* felt it had to remind readers the epidemic was not just a "fire drill" for a terrorist attack but a trial run for a future pandemic.

Then the meaning shifted. As sobriety about the war on terror set in around the middle of the decade, the same phrase started being used to downplay or dismiss the real and growing risks of bioterrorism and other human-engineered threats. His examples are specific and named. During the last decade's debates over gain-of-function avian influenza experiments, proponents argued that since nature is the real bioterrorist, the disputed benefits outweighed the potential harms. Samuel Stanley, chair of the National Science Advisory Board for Biosecurity, said in 2017:

"I believe nature is the ultimate bioterrorist and we need to do all we can to stay one step ahead." And Twist Bioscience CEO Emily Leproust, who has written at length about how gene synthesis companies can guard against misuse, also invoked the phrase to downplay such risks, saying the largest risk is from nature rather than from some postdoc or high schooler.

His biological argument is that nature is a bad optimiser for harm. Nature has no interests, motives or political goals. To the extent it can be said to want anything, it wants to enhance populations' differential reproductive success, which only rarely aligns with hurting humans. The trillions of bacteria in the average human colon have adapted toward peaceful and often mutually beneficial coexistence, and even deadly pathogens may evolve to make humans less sick if that opens up more transmission. Natural selection is also constrained: like human engineers it must work around physiological trade-offs between traits, such as genome replication rate versus mutation rate, but unlike them it is near-sighted, improving fitness in the current environment with no attention to others.

He is careful not to overclaim. Trained biologists, let alone terrorists, still have difficulty one-upping natural selection, and our understanding of biology is very much in its infancy. But capabilities are maturing fast,

and he adds two risk sources beyond terrorism: states, which have vastly more resources (about 23 are known or suspected to have maintained biological weapons programmes in the 20th century, and some, like North Korea's, likely persist), and accidental releases, with dozens of dangerous agents and toxins reported lost or released from US labs every year.

Why it's assigned

It is the first Core reading on the 1.3 page and it does the groundwork for the entire unit: it clears away the one objection that would otherwise make the rest of the day unnecessary. If nature is already the worst case, there is nothing to prepare for beyond ordinary public health. Bakerlee's answer, that nature sets no ceiling because it has no intent, is the premise that 1.3 #4, #5, #7 and #8 all build on. It is also a short, well-aimed piece of rhetorical hygiene: the course wants her to notice when a phrase is doing an argument's work for it.

In depth

The phrase and where it came from

Bakerlee opens by granting the point his opponents start from: emerging pathogens have a menacing track record. What he objects to is the slogan built on that track record. The saying was born after September 11, 2001 and the anthrax attacks, and its

purpose then was a warning against **fixating on bioterrorism** while neglecting natural pathogens.

He says that warning was right. In the following years the all-consuming war on terror fuelled a tenfold boost in biodefense spending, and the focus on bioterrorism grew so strong that when SARS hit the world stage in 2003 the editorial board of the journal *Nature* felt the need to remind its audience that the pandemic was not just a fire drill for a terrorist attack, but also a trial run for a future pandemic.

How the meaning flipped

Sobriety and disillusionment about the war on terror set in toward the middle of the decade. In concert with that shift in public opinion, the meaning of the refrain changed. Instead of telling people to pay due attention to naturally emerging pathogens, the phrase was increasingly used to downplay or dismiss the real and growing risks posed by bioterrorism and other human-engineered threats.

Bakerlee gives three pieces of evidence for the new usage.

First, the gain-of-function debates. Over the last decade scientists sought to study the emergence potential of deadly avian influenza by artificially enhancing its transmissibility between mammals. Proponents of what he calls unusually risky research suggested that, since nature is the real bioterrorist,

the experiments' disputed benefits outweighed their potential harms.

Second, Samuel Stanley, chair of the National Science Advisory Board for Biosecurity, who affirmed in 2017: "I believe nature is the ultimate bioterrorist and we need to do all we can to stay one step ahead."

Third, leading biotechnologists. Emily Leproust, CEO of Twist Bioscience, has written at length about how gene synthesis companies like Twist can better guard against the misuse of their technology by malevolent or reckless actors, and has also invoked the ultimate-bioterrorist meme to downplay such risks. In a published interview she said: "I am concerned with the risks. There is potential for massive loss of life. But the risk isn't from some postdoc or high schooler. The largest risk is from nature. ... Nature is the greatest bioterrorist. The biggest losses of life have been from nature."

His summary of these three: the meme no longer serves us well. It is undoubtedly a mistake to underestimate natural pathogens, and it is equally unwise to wield a 19-year-old expression like a magic wand to briskly banish concerns about people causing harm with biology. We cannot afford to blind ourselves to the uncomfortable truth that humans grow more capable, with each passing day, of outdoing nature and harnessing biotechnology to cause harm on a staggering scale, by either cruelty or carelessness.

Why nature is a poor optimiser for harm

This is the substantive core. Nature has no interests, motives or political goals. To the extent it can be said to want anything, it is to perpetually enhance populations' differential reproductive success, which only rarely aligns with causing greater harm to humans.

Two illustrations. The trillions of bacteria living in the average human's colon appear to have adapted toward a peaceful and often mutually beneficial coexistence with their host. And even deadly pathogens may theoretically evolve toward making humans less sick, if doing so opens up more opportunities for transmission between hosts.

Natural selection is also constrained in two distinct ways. Like human bioengineers, it must work around stubborn physiological trade-offs between traits, such as genome replication rate and mutation rate. But it is additionally handicapped by near-sightedness: it drives improvements in traits that enhance a population's fitness in its current environment, with no attention to maintaining or improving traits that would help in other environments. The process, for all its power, is therefore highly constrained in its ability to generate "superbugs" possessing a diabolical suite of traits.

The figure that is not a figure

Bakerlee's analogy does the work a chart would. If creating an especially deadly pathogen were like winning a soccer match against a formidable opponent, natural selection would be competing with all the cunning of an especially persistent horde of five-year-olds, glued to the ball and only ever capable of playing offense, defense or goalie at any one time. Modern biologists, by contrast, are gaining the ability to see the whole field, develop an intuition about where the ball will be next, and play multiple positions simultaneously.

The point is that intent plus foresight beats blind local optimisation. Through a combination of rational design, directed evolution, breeding and brute force trial and error, biologists can increasingly engineer organisms that excel at several desired functions at once, such as growing quickly in a massive industrial fermenter while churning out commercially valuable biomolecules. He is explicit that this capability promises tremendous benefits for agriculture, industry and human health, and that its potential application to pathogens is what poses serious concerns.

What he does not claim

He emphasises that trained biologists, let alone terrorists, still have difficulty one-upping natural selection's creative output, and that our understanding of biology is very much in its infancy. His claim is about trajectory: knowledge and

capabilities are maturing rapidly, as shown by Twist's prolific gene synthesis capabilities and recent feats in predicting protein structure, gene editing and genome assembly. We are much closer to this exciting but frightening horizon today than in 2001, and the trend will likely persist.

States and accidents

Bakerlee then widens the risk beyond the bioterrorist the phrase argues about. When it comes to weapons-grade biotechnology, states likely pose a greater risk than non-state terrorists, because they have vastly more resources. About 23 states are known or suspected to have maintained biological weapons programmes in the 20th century, and some programmes, like North Korea's, likely persist to this day. As countries jockey for advantage, state programmes remain an ever-present danger despite the treaties and export controls designed to rein them in, and COVID-19, which exposed countries' vulnerability to biological threats, has done little to mitigate that danger.

Accidental releases are a third source. Thanks to the US government's monitoring programme, we know that dozens of agents and toxins with the potential to pose a severe threat to public health and agriculture are reported accidentally lost or released from US labs every year. Accidental releases around the world have already caused significant harm, and such risks

increase as biotechnology expands across the world and gains in strength.

The close

Biotechnology, with all its promise and peril, is moving fast. It is irresponsible, he writes, to shrug off current and emerging biotechnological threats by reciting "Nature is the ultimate bioterrorist" like an article of faith. He compares it to global warming: the cost of wilful ignorance and inaction is high and increasing. Health security requires engaging cautiously but honestly with the full spectrum of evolving biological risks, striving toward solutions with open eyes and moral courage.

The byline matters for reading the piece: Bakerlee was then a PhD candidate studying evolutionary genetics at Harvard University and a fellow in the Council on Strategic Risks's Fellowship for Ending Bioweapons Programs. The evolutionary-biology argument is his own field.

Words to know

- **Gain-of-function research** — experiments that give an organism a new or enhanced property, here making avian influenza more transmissible between mammals. The phrase Bakerlee is attacking was used to justify exactly this work.

- **Differential reproductive success** — the only thing natural selection can be said to optimise for: some variants leaving more descendants than others. Naming it is how Bakerlee shows that harm to humans is incidental to evolution rather than its aim.
- **Directed evolution** — deliberately applying selection pressure in the laboratory to get a trait you want. It is one of the tools that lets human designers do what blind selection cannot.
- **Rational design** — engineering a biological molecule or organism from an understanding of how it works, rather than by trial and error. Together with directed evolution it is the "seeing the whole field" half of his analogy.
- **National Science Advisory Board for Biosecurity (NSABB)** — the US federal advisory body on dual-use life-science research. Its chair is quoted here using the phrase, which is Bakerlee's evidence that this is not a fringe habit.
- **Export controls** — legal restrictions on transferring sensitive goods and technology across borders, named as one of the tools that has not stopped state programmes.

Numbers in this reading

- The phrase is 19 years old at the time of writing (traced to 2001 and its aftermath).

- A tenfold boost in US biodefense spending in the years after 2001.
- About 23 states are known or suspected to have maintained biological weapons programmes in the 20th century.
- Dozens of agents and toxins with potential to pose a severe threat to public health and agriculture are reported accidentally lost or released from US labs every year.
- Dates that anchor the argument: September 11, 2001; the 2003 *Nature* editorial on SARS; Stanley's statement in 2017; publication 8 January 2021.

What the author is unsure about

Bakerlee concedes the strongest version of the other side: trained biologists, let alone terrorists, still have difficulty one-upping natural selection, and our understanding of biology is in its infancy. His claim that deadly pathogens may evolve toward making humans less sick is marked as theoretical. His statement about North Korea's programme is hedged as "likely" persisting, and the 23 states are "known or suspected". His central claim is explicitly about direction rather than level: we are closer to the horizon than in 2001 and the trend will likely persist. What he does not hedge is the rhetorical judgement, which is that the phrase has stopped serving us and should be retired, and that treating it as an article of faith is irresponsible.

For the discussion

- A point worth making: the piece's real move is about intent, not capability. Nature has no interests, motives or political goals, so its track record sets a floor on harm and tells us nothing about a ceiling.
- A tension worth raising: Bakerlee names a gene synthesis CEO using this phrase while also writing seriously about screening for misuse. The same people can hold both positions, which suggests the slogan is doing something other than expressing a belief about risk.

1.3-3 · FURTHER

AI and the Evolution of Biological National Security Risks

Bill Drexel and Caleb Withers · 2024

[Read the original paper](#) → · the course assigns: the executive summary and introduction as published on the CNAS page; the full report PDF is not in the saved text

CNAS's verdict on the AI-bio alarm is that the theoretical potential is very large but the measured impact so far has been marginal, so the right response is to monitor four specific capabilities rather than act on predictions.

Average summary

From the executive summary and introduction published on the CNAS page; the saved text ends with a "Read the full report" link, so the body of the report is not in the capture.

Drexel and Withers open with the claims they are testing. The CEOs of leading AI labs have raised alarms about new technical capabilities for would-be bioterrorists; Dario Amodei told a July 2023 congressional hearing that within two to three years there was a substantial risk AI tools would greatly widen the range of actors with the technical capability to conduct a large-scale biological attack; Rishi Sunak spoke of a small window before AI enables a step change in bioterrorist capability; and Kamala Harris warned of AI-formulated bioweapons that could endanger the very existence of humanity. The report's question is blunt: are they true?

Its answer is layered. On the pre-AI baseline, policymakers have historically underappreciated the risks from the routine activities of well-intentioned scientists, even as the number of high-risk biosecurity labs and the frequency of dangerous incidents (perhaps including COVID-19 itself) continue to grow. States have traditionally been a major source of risk, not least the Soviet programme, but the unwieldiness and imprecision of bioweapons means they remain unlikely to field large-scale attacks in the near term. Non-state actors have an unnerving track record of

attempting attacks but with limited success, because of the intrinsic complexity of building and wielding such delicate capabilities. Biotechnology is changing that baseline independently of AI, and cloud labs could let non-state actors outsource the experimental expertise that has historically been the barrier.

Experts disagree: the less concerned point to the practicalities of constructing, storing and disseminating a weapon and to the absence of a major attack, while the more pessimistic point to experiments showing powerful viruses built from commercially available inputs. On AI specifically, the most pressing concern is not novel design but acceleration of procurement by non-state actors, and in particular AI's budding ability to help **troubleshoot where experiments have gone wrong**, speeding the design-build-test-learn feedback loop. More speculative capabilities, such as optimising a weapon for specific genetic groups or geographies, or creating superviruses optimised for transmissibility and lethality at once, would dramatically alter state incentives if realised, but face technical challenges that may not be solvable.

Because predicting timelines is, in their phrase, a fool's errand, they name four capability areas to watch: foundation models' ability to provide experimental instructions for advanced biological applications; cloud labs' and lab automation's progress in reducing

the need for experimental expertise; dual-use progress in research on host genetic susceptibility to infectious diseases; and dual-use progress in precision engineering of viral pathogens. Their five recommendations are to strengthen screening for cloud labs and gene synthesis providers, assess foundation models across the full bioweapons lifecycle regularly, invest in technical safety mechanisms, make biodefense investment more agile, and in the long term consider a licensing regime for a narrow set of biological design tools, if catastrophic capabilities begin to materialise.

Why it's assigned

It is the Further-tier policy report on the 1.3 page, and it is the most disciplined answer to the day's second question. Where the other readings argue that the ceiling on harm is rising, this one asks what evidence would show that, and answers with four watchable indicators and a set of actions that do not depend on a forecast being right. It also supplies the quoted judgement that 1.3 #1 leans on about states being unlikely to attack while non-state actors keep trying.

Words to know

- **Foundation model** — a large general-purpose AI model that many applications are built on. The report's central concern is what these can and cannot do across a bioweapons lifecycle.

- **Cloud lab** — a facility that runs experiments on a client's behalf. It matters because experimental expertise, not knowledge, has been the main barrier for non-state actors.
- **Design-build-test-learn loop** — the iterative cycle by which biological work actually gets done. The report's sharpest claim is that AI's real contribution is speeding this loop through troubleshooting, not inventing anything new.
- **Dual-use** — research or technology with both beneficial and harmful applications. Two of the four capabilities to monitor are explicitly dual-use research areas.
- **Unlearning** — a technical safety approach that removes specific knowledge from a trained model. Named as an investment priority.
- **Information hazard** — information that causes harm simply by existing or being disclosed. The report asks for novel approaches to this problem in model training.
- **Host genetic susceptibility** — research into why particular people or populations are more vulnerable to a given infection. Legitimate medical research, and also the route to the targeted weapons the report calls speculative.

Numbers in this reading

- COVID-19: nearly 29 million estimated deaths.

- Amodei's July 2023 testimony: within two to three years, a substantial risk AI tools would greatly widen the range of actors capable of a large-scale biological attack.
- Four capability areas to monitor, and five recommended measures.
- Scaled historical comparisons in the notes: the 1918 influenza killed approximately 1 to 2 percent of the world's population, equivalent to 70 to 150 million today; the Black Death killed about half of Europeans over a few years in the mid-1300s.
- Jeff Alstott, former director for technology and national security at the National Security Council, warned in September 2023 that the classified record contained "fairly recent close-ish calls" of non-state actors attempting to use biological weapons with strategic scale.

What the author is unsure about

This report is mostly made of calibrated uncertainty. The authors say that to date AI's impacts on biological risks have been marginal, and that there is no way to know for certain if or when more severe risks will materialise. They describe the targeting and supervirus capabilities as speculative, and note technical challenges scientists would need to overcome, "if they can be solved at all". They present the optimistic and pessimistic expert camps without picking one, and they say accurately weighing how

credible the dire warnings are "can feel next to impossible" and requires judgement in the face of opaque counterfactuals, alarmism, denialism and horrific possibilities. They call predicting exactly how and when these risks will develop a fool's errand, given inherent uncertainty in both disciplines. Their licensing recommendation is explicitly conditional on capabilities beginning to materialise. The one thing they describe as less controversial is the insufficiency of US biodefences.

For the discussion

- A point worth making: the report reframes the question usefully. The threat is not that a model designs a new pathogen, it is that a model shortens the loop between a failed experiment and a working one, which is where tacit knowledge used to protect us.
- A tension worth raising: the authors say AI's impact so far has been marginal and also recommend regular rigorous evaluations across the whole bioweapons lifecycle. That is a bet on measurement over prediction, and it only works if the evaluations are good enough to catch a threshold being crossed.
- A question worth asking: two of their four watch items are ordinary medical research areas, host genetic susceptibility and precision viral

engineering. Who is supposed to be monitoring those, and with what authority?

1.3-4 · CORE · COURSE TIME 8 MIN

Nature Sets a Baseline, Not a Ceiling, for Biological Harm

Becca Chandler Bostock · 2026

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

Bostock, writing from SecureBio, picks up Bakerlee's 2021 argument five years on and adds the thing that has changed: AI is reducing the expertise and luck deliberate misuse requires, so engineered threats now deserve at least equal billing with natural ones.

Average summary

The piece starts from field experience rather than theory. In more than 30 conversations with academic researchers about the risks of AI-enabled biological threats, Bostock says one objection kept recurring: that **nature is the worst bioterrorist**. The phrase is often attributed to Anthony Fauci, who said in a 2018 CBS interview that nature is potentially the worst bioterrorist for us, but it has circulated since at least a 2003 *Nature* editorial that called the emerging SARS

epidemic a timely reminder that Mother Nature is the ultimate bioterrorist.

She concedes the historical point: engineered biological agents have been poor terrorist weapons compared with alternatives. Her claim is about 2026, when AI is greatly reducing the expertise and luck deliberate misuse requires, so it should be considered at least an equally serious threat alongside naturally occurring pathogens. She credits Bakerlee's 2021 piece and adds that in 2025 the Nuclear Threat Initiative issued a statement signed by more than 40 experts warning that AI-enabled tools could make it possible to design pathogens more dangerous than what is found in nature.

Humans can scale natural processes. Nature supplies the parts and humans assemble them for outcomes nature never produces, mostly beneficially: bacteria engineered to make insulin at industrial scale have a capability they would never evolve. But scale cuts both ways. Inhalational anthrax is vanishingly rare in nature even though *Bacillus anthracis* lives in soil worldwide; in 1979 a Soviet spore production facility in Sverdlovsk suffered a filter failure and released a plume across the city, killing at least 66 people across an area up to 4 km downwind.

Nature has no intent. Casting nature as a bioterrorist highlights the dangers of natural infection rather than claiming nature has an agenda, but the metaphor

muddies the water by linking the two. Nature optimises for reproductive fitness: replication, transmission, persistence. Classical theory holds there is a trade-off between transmission and virulence, settling viruses at intermediate virulence, which is why natural pathogens rarely become both highly transmissible and highly lethal. But there is no fundamental law of nature saying a pathogen cannot be both. An engineered pathogen would face the same evolutionary pressures once released, but that takes time, and acute harm happens fast. A malicious actor can also choose where and when to release, raising the ceiling of harm without any novel biology.

Where AI fits. AI cannot yet design a novel pathogen from scratch, but general-purpose models provide scientific expertise and troubleshooting, and biological AI models (BAIMs) specialised on biological sequences, structures or experimental data can predict molecular structures and interactions and estimate regulatory activity and variant effects. She names examples: AlphaFold 3, ESM3, AlphaGenome, Evo 2. These tools screen many variants in silico and flag the most promising for laboratory work, compressing the time to optimise biological variants and narrowing the window for law-enforcement intervention. Evo models have generated viable Φ X174 bacteriophage genomes not previously detected in nature, though researchers dispute how novel they are.

Why it's assigned

This is the 2026 update to the Core argument 1.3 #2 makes, and the course pairs them deliberately: read the original retirement notice for the phrase, then read what five years of AI capability did to the question. It is also the reading that introduces the vocabulary the rest of the week uses for AI-bio tools, and the one that makes the course's own case for pathogen-agnostic detection, which becomes Unit 3's subject.

In depth

The objection she keeps hearing

Bostock's framing is empirical. Across more than 30 conversations with academic researchers about AI-enabled biological threats, one objection recurred as a reason not to take deliberate misuse seriously: **nature is the worst bioterrorist**. A footnote adds that colleagues at SecureBio heard the same thing in their own conversations.

She traces the phrase. It is often attributed to Anthony Fauci, who said in a 2018 CBS interview that nature is potentially the worst bioterrorist for us. But it has been in public discourse since at least a 2003 editorial in *Nature* calling the emerging SARS epidemic a timely reminder that Mother Nature is the ultimate bioterrorist. A second footnote pushes the reasoning back further: an editorial from 2000 called

for scientists to be careful not to exaggerate the usefulness of biological agents as terrorist weapons, and notes the political context of that threat environment, including ongoing Biological Weapons Convention verification negotiations and shifting funding priorities, with both pharmaceutical companies and the US military and related industries having stakes in the game.

Her own position is a concession plus a timestamp. She agrees that historically, engineered biological agents have been poor terrorist weapons compared with alternatives. In 2026, however, AI is greatly reducing the expertise and luck required for deliberate misuse, and should be considered at least an equally serious threat alongside naturally occurring pathogens.

She places herself in a lineage: Bakerlee argued in 2021 that the phrase had strayed from its purpose of warning against neglecting natural pandemics and become a way of discounting engineered threats, his key point being that nature has no interests, motives or political goals. In 2025 the Nuclear Threat Initiative issued a statement signed by more than 40 experts warning that AI-enabled tools could make it possible to design pathogens more dangerous than what is found in nature.

Humans can scale natural processes

As humans we already harness and exceed natural biological processes. Nature supplies the parts; humans assemble them to achieve outcomes not found in the natural world. For the most part this is hugely beneficial. Her example is bacteria engineered to produce insulin on an industrial scale: by engineering them, scientists give bacteria a capability they would never evolve, and manufacture huge quantities of highly purified insulin, which lets people with diabetes manage their condition with greater ease.

Producing biological products or agents at industrial scale can also cause harm, and her example is exact. Anthrax inhalation is vanishingly rare in nature, even though *Bacillus anthracis* lives in soil across the globe. In 1979, a Soviet spore production facility in Sverdlovsk suffered a filter failure and released a plume of spores across the city. The resulting outbreak of inhalational anthrax killed at least 66 people across an area up to 4 km downwind. The point is that human decisions about cultivation and distribution, not any change to the organism, produced a form of harm nature essentially never produces.

Nature has no intent

Casting nature as a bioterrorist, she says, highlights the dangers of natural infection rather than suggesting nature has an agenda or intentionality as a

human bioterrorist would. Nevertheless the phrase muddies the water by metaphorically linking the two, and much of the concern about engineered threats stems directly from the fact that they would be created with the intention of harm.

Nature optimises for reproductive fitness: replication, transmission and persistence. The classical theory is that there is a trade-off between transmission and virulence, so viruses settle at an end-point of intermediate virulence, which is why natural pathogens rarely evolve to be both highly transmissible and highly lethal. Her footnote qualifies this honestly: the trade-off is real but perhaps not universal, since some systems do not conform to classical trade-off theory, as argued for pandemic SARS-CoV-2.

Her key sentence: there is no fundamental law of nature that says a pathogen cannot be both. She grants that once released, an engineered pathogen faces the same evolutionary pressures as a natural one, and the footnote gives the textbook case: myxoma virus, a South American rabbit poxvirus causing mild skin lesions in native species, was released in Australia in 1950 to control European rabbits, caused over 99% lethality within two weeks of infection, then attenuated within a few years because the more lethal strains killed rabbits before

mosquitoes could spread them. But that takes time, and acute harm happens fast.

A second lever needs no novel biology at all: a malicious actor can choose where and when to release a pathogen to maximise panic or economic disruption, and as the anthrax example shows, human changes in cultivation and distribution can dramatically increase the realised impact of even natural pathogens.

On likelihood she is careful. The next outbreak is more likely to come from a known pathogen, such as measles in an under-vaccinated population or a variation of last year's seasonal flu, than from an engineered release, and larger outbreaks may come from novel zoonotic spillover, made more likely by deforestation and wildlife trafficking. Her reply is one line: this may not be as bad as it could get.

Where AI fits in

AI is not yet able to design a novel pathogen from scratch, but capabilities are rapidly increasing. General-purpose AI models can provide scientific expertise and troubleshooting, measured through evaluations of frontier models such as those carried out by SecureBio.

She then introduces the category that matters. Biological AI models (BAIMs) are specialised machine-learning systems trained on biological

sequences, structures or experimental data. They can predict molecular structures and interactions and estimate regulatory activity and variant effects. Her four named examples: biomolecular structure predictors such as AlphaFold 3; protein language models such as ESM3; DNA sequence-to-function models such as AlphaGenome; and genome language models such as Evo 2.

What they do for research is speed it up: they screen many possible variants *in silico* and highlight the most promising for laboratory verification, saving the time and resources scientists would need to test every variant in the laboratory. Her worked example: Evo models have generated viable Φ X174 bacteriophage genomes (a virus that infects bacteria) that were not previously detected in nature. She immediately adds the caveat that researchers dispute the extent of their novelty, and then states what the work does demonstrate regardless, which is that BAIMs can compress the time it takes scientists to optimise biological variants, narrowing the window for law-enforcement intervention. These tools are also opening up entirely new protein variant possibilities, and BAIMs together with more traditional laboratory methods can generate proteins that would not arise in nature.

Then the argument she most wants to rebut. AI tools still fail at some tasks, such as predicting viral

epistasis or designing a completely novel virus. She has heard this offered as proof that we cannot build anything worse than nature. Her reply is a model of careful reasoning: it may in fact be the case that we cannot now build anything worse than nature. But if success at such tasks would demonstrate the latent potential for a dangerous use, should we not keep an eye on them? That is what makes it worth spending time and resources measuring AI capabilities and putting safeguards in place before these thresholds are crossed.

She closes the section on intent and capability converging. Boko Haram used AI to plan attacks in Nigeria in 2014, and al-Qaeda ran a biological weapons programme focused on anthrax production from 1999 to 2001. The convergence of a longstanding interest in biological weapons with potential capability uplift from AI is, in her word, worrying.

Future possibilities of harm

There is huge potential for positive medical developments: improved vaccines, antivirals and antibiotics. But these technologies are often dual-use, a concern highlighted by the National Academies and the NTI among others, and she gives two symmetrical examples.

A model that predicts which virus strains will not be stopped by current vaccines can be used to develop more effective vaccines, or to design vaccine-resistant pathogens. This is the same AI capability with very different end goals.

Interspecies transmission prediction helps us prepare for pandemics by predicting which genome mutations and changes in protein structure allow a pathogen to infect new species. The same capability could allow someone to engineer an animal pathogen to spread through human populations.

Defence against biological threats

Her prescription has two halves. First, identify and reduce emerging risks: SecureBio AI's work focuses on developing evaluations to measure frontier AI capabilities, because this empirical measurement is necessary to understand where capabilities and risks currently lie, and that is vital information as society intentionally decides how to develop AI models for public good and to mitigate harm.

Second, broaden surveillance. Pandemic preparedness and outbreak surveillance are important and will remain so, but to be truly robust, surveillance and defensive efforts should include engineered pathogens. Pathogen-agnostic surveillance systems, such as those developed by SecureBio Detection, identify novel and known pathogens from

environmental samples and swab sampling, so the same work alerts us to natural outbreaks and engineered pathogens alike. She quotes a conclusion from an article in *Lancet Infectious Diseases*: preparedness for intentional outbreaks has the important added value of strengthening preparedness for natural epidemics, and vice versa.

Her final paragraph is the title restated. Natural pathogen outbreaks are real, happen frequently and can be devastating, but they are not the ceiling of possible harm. We may face worse threats in future, particularly with advances in AI capability. The argument is not that one threat matters more than the other; it is that our attention and resources should consider both.

Words to know

- **BAIM (biological AI model)** — a machine-learning system trained on biological sequences, structures or experimental data rather than on language. The distinction from a general chatbot is the whole reason this reading exists.
- **Genome language model** — a model trained on DNA sequences the way a language model is trained on text; Evo 2 is the example. It is what generated the bacteriophage genomes she describes.

- **In silico** — done by computer simulation rather than in a laboratory. The time saving from screening variants in silico is the mechanism of concern.
- **Epistasis** — where the effect of a mutation depends on the rest of the genome it sits in, rather than having a fixed effect that adds up with other mutations. Mutations in one part of a viral genome can have hard-to-predict effects elsewhere. It is currently a thing AI does badly, and she argues that is a reason to watch it, not to relax.
- **Transmission-virulence trade-off** — the classical idea that a pathogen cannot maximise both spread and lethality, because killing the host too fast prevents transmission. Her point is that this is a tendency of evolution, not a law of physics.
- **Pathogen-agnostic surveillance** — detection that looks for anything unusual rather than for a named list of pathogens. It is the defence that covers natural and engineered threats with one system.
- **Dual-use** — the same capability serving benefit and harm. Her two vaccine and interspecies examples are the cleanest illustrations in the unit.

Numbers in this reading

- More than 30 conversations with academic researchers in which the objection recurred.
- The phrase traced to a 2018 Fauci interview, a 2003 *Nature* editorial and reasoning going back to

a 2000 editorial.

- 2025: an NTI statement signed by more than 40 experts warning AI-enabled tools could make it possible to design pathogens more dangerous than those found in nature.
- Sverdlovsk, 1979: at least 66 people killed, across an area up to 4 km downwind, after a filter failure.
- Myxoma virus in Australia, 1950: over 99% lethality within two weeks of infection in European rabbits, attenuating within a few years.
- al-Qaeda's biological weapons programme focused on anthrax production, 1999 to 2001; Boko Haram used AI to plan attacks in Nigeria in 2024.
- Named models: AlphaFold 3, ESM3, AlphaGenome, Evo 2.

What the author is unsure about

Bostock is careful in both directions and says so. She agrees engineered agents have historically been poor terrorist weapons. She agrees the next outbreak is more likely to be a known pathogen or a zoonotic spillover than a deliberate release. She states that AI cannot yet design a novel pathogen from scratch. She notes researchers dispute how novel the AI-generated bacteriophage genomes actually are. She footnotes that the transmission-virulence trade-off is real but perhaps not universal. Most strikingly, she concedes the core of the counter-argument outright: it may in fact be the case that we cannot now build anything

worse than nature. Her position is that this is a reason to measure capabilities and set safeguards before thresholds are crossed, not a reason to stop watching. The word she uses for the convergence of terrorist interest and AI uplift is "worrying", which is a judgement rather than a finding.

For the discussion

- A point worth making: the Sverdlovsk example is the cleanest version of the title. Nobody edited the anthrax bacterium; a human decision to grow and concentrate it at industrial scale produced a kind of harm that nature almost never produces.
- A tension worth raising: she grants that a released engineered pathogen would attenuate under the same evolutionary pressures as any other, and then says acute harm happens before evolution catches up. That makes the argument about timescale, which is testable, rather than about biology being limitless.
- A question worth asking: her reasoning is that if a capability would prove latent potential for dangerous use, we should watch it even while models fail at it. Where does that logic stop, given that half the capabilities she names are also how better vaccines get made?

How AI could enable catastrophic pandemics

Will Saunter · 2025

[Read the original paper →](#) · the course assigns: whole piece (the post states it was updated in Nov 2025 to include the release of Claude Opus 4.5)

Saunter's state-of-play argument is that biology is civilisation's greatest vulnerability, that frontier models are on the cusp of helping novices build biological weapons while AI biological tools raise the ceiling on what is possible at all, and that the window to build defences is closing.

Average summary

This is BlueDot's own briefing, covering the state of play as of September 2025 and updated in November 2025 for Claude Opus 4.5. It opens with three claims. Frontier US large language models are "on the cusp" of being able to help novices build biological weapons, and have been released with enhanced safeguards. Chinese models show comparable capabilities with dramatically weaker safeguards, complying with **94% of malicious requests** versus 8% for US models. AI-enabled biological tools are raising the ceiling, with recent work generating 16 novel viruses, several more infectious than their natural templates, and many high-risk tools fully open-source.

The scale argument comes first. COVID-19 killed an estimated 27 million people worldwide and cost tens of trillions of dollars. Since 1900, only seven acute disasters have killed more than 10 million people, and three of them are pandemics: the 1918 influenza pandemic (50 to 100 million), the Second World War (around 66 million), the Great Chinese Famine (around 36 million), HIV/AIDS (around 33 million), COVID-19 (around 27 million), the Chinese Famine of 1906 to 1907 (20 to 25 million) and the First World War (around 15 million). For comparison, all other wars from 1800 to 2011 combined killed roughly 9 million combatants.

Synthetic biology has moved fast. Sequencing DNA is 1 million times cheaper in 2025 than in 2003; synthesis costs have fallen less, only 20 times cheaper than in 2005. The first fully synthetic virus was built in 2002 and the first fully synthetic bacterium in 2010. The benefits are real (synthetic human insulin since 1983, semi-synthetic artemisinin, mRNA COVID vaccines developed 10 times faster than any vaccine in history) and so are the new weapon categories: gene drives, and mirror life, where leading synthetic biologists have urged caution.

His worked example is Evo 2, a genome language model trained on DNA from over 100,000 species. In September 2025 researchers used it to build 16 synthetic viruses based on Φ X174, a bacteriophage

targeting *E. coli*; many are distinct enough to be classified as their own species and some were considerably more infectious than the original. The Evo 2 team excluded human-infecting viruses from the training data, but the code, model parameters and training data are fully open source, and Saunter notes that two minutes of searching found a public fine-tuning tutorial and a large US government open-source database of human viral genomes.

He ends on why no attack has happened yet (tacit knowledge, wet lab skills, equipment access, a small pool of actors with both capability and intent, and models with meaningful assistance capabilities existing for less than a year) and on why that is not reassuring: the absence of attacks tells us more about past barriers than about future risk. His prescription names DNA synthesis screening as the critical chokepoint between design and physical production, which should be mandatory and global, alongside metagenomic sequencing, PPE stockpiles and mRNA platforms.

Why it's assigned

This is the course's own answer to the day's second question, written by BlueDot, and it is the reading that stitches the unit together: it supplies the comparative death tolls that 1.2 established, the engineered-ceiling argument that 1.3 #2 and #4 make, the mirror life problem that 1.3 #9 covers, and the chokepoint logic

that becomes Unit 2 (prevent) and the detection, PPE and countermeasure stack that becomes Units 3, 4 and 5. If she reads one piece to understand why this course exists, it is this one.

In depth

Biology is human civilisation's greatest vulnerability

Saunter's opening move is comparative. COVID-19 killed an estimated 27 million people worldwide and cost tens of trillions of dollars, continuing a centuries-long trend of pandemic disease being among humanity's deadliest disasters.

He makes that concrete with a list of **only seven acute disasters since 1900 that killed more than 10 million people**: the 1918 "Spanish Flu" pandemic (1918 to 1920), 50 to 100 million deaths; the Second World War (1939 to 1945), around 66 million; the Great Chinese Famine (1959 to 1961), around 36 million; the HIV/AIDS pandemic (1981 to present), around 33 million; the COVID-19 pandemic (2019 to 2024), around 27 million; the Chinese Famine of 1906 to 1907, 20 to 25 million; and the First World War (1914 to 1918), around 15 million. Three of the seven are pandemics.

The two comparisons he sets against that list are the part worth remembering. All other wars from 1800 to 2011 combined killed roughly 9 million combatants,

and all non-disease and non-famine natural disasters since 1900 combined killed roughly 23 million people. A self-replicating, invisible pathogen spreading from human to human is a devastating way of killing many people fast.

This has not escaped governments, terrorist groups and lone rogue actors. Through the 20th century the Soviet Union, Japan, the US and many other states had major bioweapons programmes, and Russia, North Korea and China are widely suspected of maintaining them today despite signing the Biological Weapons Convention. The Rajneeshee and Aum Shinrikyo groups both attempted biological attacks in the 1980s and 90s, Aum Shinrikyo with explicitly apocalyptic aims, and the 2001 anthrax attacks were likely perpetrated by the lone wolf attacker Bruce Edwards Ivins, a scientist at the US government's biodefense labs at Fort Detrick.

Synthetic biology

Ever since the mid-1900s discovery that DNA is the primary hereditary material, the race has been on to understand and manipulate it, and the first crude tools to read and edit DNA arrived in the 1970s and 80s.

His cost numbers are worth keeping straight. Sequencing DNA is 1 million times cheaper in 2025 than in 2003, when the Human Genome Project was

completed, while synthesising DNA is only 20 times cheaper today than in 2005. The first DNA cutting enzymes of the 1970s could only cut specific sites, whereas today's CRISPR systems edit with near single-nucleotide precision. We can also write and print genomes from scratch: the first fully synthetic virus was built in 2002 and the first fully synthetic bacterium in 2010, and since then we have built novel bacteria with all unnecessary genes deleted, and recoded genomes.

The benefits are real: synthetic human insulin produced by bacteria has transformed diabetes care since 1983; semi-synthetic artemisinin has stabilised malaria drug supply and reduced price volatility; and mRNA COVID-19 vaccines were developed and licensed 10 times faster than any other vaccine in history, likely saving millions of lives.

Then the other side, with two examples. Gene drives are DNA sequences that replicate themselves so they pass to all offspring in sexually reproducing organisms; proposed to make mosquito populations resistant to the malaria parasite and effective in lab tests, they could in theory be designed for crops or even humans. Mirror life means organisms whose DNA and proteins are the mirror image of those in nature, still decades from being possible; the concern is that mirror bacteria could feed on non-chiral nutrients, which are symmetrical so the mirror image

is the same as the natural version, could infect humans, animals and crops, and our immune systems would struggle to recognise and kill them.

He quotes the mirror life paper at length, and it is the most striking passage in the piece because it records researchers changing their minds: initially sceptical, they became deeply concerned, concluded technological progress will likely make mirror bacteria possible, and realised that generalist heterotroph mirror bacteria would not be intrinsically biocontained, so mirror organisms should not be created absent compelling evidence for reassurance. (The same report is the subject of 1.3 #9.) Saunter's gloss is modest: the example is still theoretical, but it shows how synthetic biology could expand the possibility space beyond what nature has made possible.

AI, part one: large language models

He first notes AI's upside: the 2024 Nobel Prize for Chemistry went to the designers of AlphaFold2, which predicts protein structure from amino acid sequences.

On risk he names two areas experts worry about: LLMs enabling more people to develop and deploy dangerous pathogens by helping novices design and troubleshoot experiments; and AI-enabled biological

tools raising the ceiling for harm through novel synthetic biology.

Anthropic and OpenAI are the only companies that publish detailed bioweapons evaluations, measuring the ability to make end-to-end plans for synthesising a biological weapon, troubleshoot wet lab experiments and evade DNA synthesis screening.

Both say they are on the cusp of being able to assist novices, and have released the models with appropriate safeguards. He quotes four releases.

OpenAI's GPT-5 (August 2025) was treated as High capability in the Biological and Chemical domain, activating Preparedness safeguards, while OpenAI stated it lacked definitive evidence the model could meaningfully help a novice create severe biological harm and that the model remains on the cusp.

Anthropic's Claude Opus 4 (May 2025) came sufficiently close that they were unable to rule out ASL-3, the threshold for significantly helping people with basic technical backgrounds, such as undergraduate STEM degrees, to create, obtain and deploy CBRN weapons. Claude Sonnet 4.5

(September 2025) showed improved biology knowledge and tool use, outperforming Claude Opus 4 and 4.1 on both long-form virology tasks, so ASL-3 safeguards were deemed appropriate. Claude Opus 4.5 (November 2025) was meaningfully more helpful in an expert uplift trial, with higher scores and fewer critical errors, but still produced critical errors yielding

non-viable protocols, and Anthropic added that the CBRN-4 rule-out is less clear than they would like, partly because of limited understanding of the threat model.

Then the point that changes the policy picture.

Chinese models such as DeepSeek V3.1, Qwen3 and Kimi K2 score comparably to US frontier models on scientific reasoning, maths and general reasoning, although detailed bioweapon evaluations are lacking, and their safeguards are much weaker. A September 2025 study by the US Center for AI Standards and Innovation found DeepSeek's most secure model (R1-0528) complied with 94% of overtly malicious requests using common jailbreaking techniques, against 8% for US reference models. His conclusion: while US labs may successfully implement safeguards, Chinese models provide an accessible alternative for potential misuse.

AI, part two: biological tools

"AI-enabled biological tools" (BTs) is his catch-all for AI models trained on biological data. They are good at one narrow task relative to the general capabilities of large language models, with AlphaFold predicting protein structure as the example.

The survey numbers are the ones to remember. The most comprehensive study to date considered 1,107 tools spanning viral vector design, protein

engineering and host-pathogen interaction prediction. Of those, 57 were state-of-the-art; of those 57, 13 were rated "red" for high misuse-relevant capabilities combined with technological maturity; and most concerning to him, 61.5% of red-designated tools are fully open-source.

The figure that is not a figure: Evo 2 and the 16 viruses

His illustrative example is the genome language model Evo 2. Rather than being trained on natural language, it is trained on DNA sequences from over 100,000 different species.

In September 2025, researchers used Evo 2 to generate the first novel functional genomes created using AI, building 16 synthetic viruses based on Φ X174, a bacteriophage that targets *E. coli*. Many are sufficiently distinct from any known virus that they would be classified as their own species, and some were considerably more infectious than the original.

The safeguard and its hole are both reported. Given biosecurity concerns, the Evo 2 team excluded viruses that infect humans from the training data. However the code, model parameters and training data are fully open source, and a sufficiently motivated actor could likely fine-tune the model on those viral sequences. Saunter adds a bracketed note in his own voice: in two minutes of searching he found a publicly available

tutorial on how to fine-tune Evo 2, and that the US Government maintains a large open-source database of human viral genomes.

The window to build defences is closing

He asks why there has been no AI-enabled bioterror attack yet, and lists the remaining barriers honestly. Building and releasing a pandemic pathogen is still difficult and requires tacit knowledge beyond what LLMs provide. Wet lab skills, equipment access and troubleshooting ability still matter. The pool of actors with both capability and intent to cause mass casualties remains small. And crucially, AI models with meaningful assistance capabilities have existed for less than a year.

But the barriers are eroding: LLMs are improving at interactive troubleshooting, biological automation is reducing hands-on skill requirements, and as capability barriers fall the pool of actors with access increases. His conclusion is the line the course keeps coming back to: the absence of attacks so far tells us more about current and historical barriers than about future risk.

His prescription names one chokepoint and three layers. Comprehensive DNA synthesis screening is the critical chokepoint between in silico design and physical production, and should be mandatory and global before AI-designed sequences become trivially

accessible. Beyond that, scale existing pandemic preparedness infrastructure: metagenomic sequencing for early outbreak detection, PPE stockpiles for immediate deployment, and mRNA platform technologies for rapid vaccine production. A lot of the required technology already exists and just needs to be funded and deployed.

Words to know

- **Gene drive** — a DNA sequence that copies itself so it passes to all offspring, spreading through a sexually reproducing population rather than half of it. Proposed for malaria mosquitoes; could in theory be designed for crops or humans.
- **Mirror life** — organisms built from the mirror-image versions of natural biological molecules. The concern is that our immune systems and natural predators would not recognise them.
- **Non-chiral nutrients** — molecules that are symmetrical, so the mirror image is identical to the natural version. This is why mirror bacteria would not starve, and so would not be self-contained.
- **ASL-3 (AI Safety Level 3)** — Anthropic's capability threshold for CBRN risk: significantly helping people with basic technical backgrounds to create, obtain and deploy such weapons. Worth knowing because the quotations turn on whether a model can be "ruled out" of it.

- **Biological AI tools (BTs)** — AI models trained on biological data that are narrow but strong at one task, as distinct from general-purpose chatbots. The survey's 1,107 tools are all of this kind.
- **Genome language model (gLM)** — a model trained on DNA sequences rather than text; Evo 2 is the example, trained on over 100,000 species.
- **Bacteriophage** — a virus that infects bacteria. ΦX174 is the template for the 16 synthetic viruses, and it is a bacterial virus, not a human one.
- **Fine-tune** — further training of an existing model on new data to specialise it. It is the step that would undo the Evo 2 team's decision to exclude human viruses.
- **Jailbreaking** — techniques that get a model to comply with requests its safeguards should refuse. The 94% versus 8% comparison is a jailbreak-resistance measurement.

Numbers in this reading

- COVID-19: an estimated 27 million deaths worldwide, tens of trillions of dollars in cost.
- The seven acute disasters since 1900 killing more than 10 million: 1918 flu 50 to 100 million; Second World War around 66 million; Great Chinese Famine around 36 million; HIV/AIDS around 33 million; COVID-19 around 27 million; 1906 to 1907 Chinese Famine 20 to 25 million; First World War around 15 million.

- All other wars 1800 to 2011 combined: roughly 9 million combatants. All non-disease and non-famine natural disasters since 1900 combined: roughly 23 million people.
- DNA sequencing is 1 million times cheaper in 2025 than in 2003. DNA synthesis is 20 times cheaper than in 2005.
- First fully synthetic virus 2002; first fully synthetic bacterium 2010; synthetic human insulin transforming diabetes care since 1983; mRNA COVID vaccines developed and licensed 10 times faster than any other vaccine in history.
- Safeguard compliance: DeepSeek R1-0528 complied with 94% of overtly malicious requests using common jailbreaking techniques, versus 8% for US reference models (US Center for AI Standards and Innovation, September 2025).
- The AI-bio tool survey: 1,107 tools considered, 57 state-of-the-art, 13 rated red, 61.5% of red tools fully open-source.
- Evo 2: trained on DNA from over 100,000 species; used in September 2025 to build 16 synthetic viruses based on ΦX174, some considerably more infectious than the original.
- AI models with meaningful assistance capabilities have existed for less than a year.

What the author is unsure about

Saunter states his limits in his own voice and lets his quoted sources state theirs. He says building and releasing a pandemic pathogen is still difficult, requires tacit knowledge beyond what LLMs provide, and that the number of actors with both capability and motivation remains small. He reports that experts still disagree on whether AI biological tools will be able to produce agents more dangerous than naturally evolved pathogens, and gives both sides without choosing. The quotations carry their own hedges: OpenAI says it does not have definitive evidence GPT-5 could meaningfully help a novice and that the model remains on the cusp; Anthropic says it is unable to rule out ASL-3 for Claude Opus 4, that Claude Opus 4.5 still produced critical errors yielding non-viable protocols, and that the CBRN-4 rule-out is less clear than they would like partly because of their limited understanding of the threat model. The mirror life authors call for scrutiny of their own findings. He notes detailed bioweapon evaluations are lacking for the Chinese models whose safeguards he criticises, and that mirror organisms are still decades from being possible. His forward-looking claims (barriers eroding, the window closing) are arguments, not measurements.

For the discussion

- A point worth making: the strongest number in the piece is not about capability at all, it is the 94%

versus 8% safeguard gap. If US labs harden their models and comparable models elsewhere comply with almost any malicious request, then safeguards at the frontier lab are a national policy, not a global one.

- A tension worth raising: the Evo 2 team deliberately excluded human-infecting viruses from training, and the model, its parameters and its training data are open source, with a public fine-tuning tutorial and a government viral genome database available. That is a case study in a safeguard that depends on everyone downstream choosing not to undo it.

1.3-6 · FURTHER

CBRN evaluations, Claude Opus 4 System Card

Anthropic · 2025

[Read the original paper →](#) · the course assigns: section 7.2

CBRN evaluations only, PDF pages 88 to 103

Anthropic publishes the actual evaluations behind the phrase "on the cusp": Claude Opus 4 could not be ruled out of needing ASL-3 safeguards, mostly on the strength of one uplift trial and two long-form virology tasks, while every participant and every model still hit critical failures.

Average summary

Read: section 7.2 only, the CBRN part of Anthropic's Responsible Scaling Policy evaluations, PDF pages 88 to 103. The rest of the system card is not covered here.

Two thresholds organise everything. The **ASL-3** capability threshold measures whether a model can significantly help individuals or groups with basic technical backgrounds, for example undergraduate STEM degrees, to create or obtain and deploy chemical, biological, radiological or nuclear weapons. The ASL-4 threat model is about substantially uplifting moderately-resourced state programmes, through novel weapons design, substantially accelerating existing processes, or dramatically reducing technical barriers. Anthropic says it primarily focuses on biological risks with the largest consequences, such as pandemics, and that it studies long multi-step tasks rather than single prompts and responses, measuring success against what could be achieved with tools available in 2023.

The results. Claude Opus 4 showed improved biology knowledge in specific areas and improved tool use for agentic biosecurity evaluations, but mixed performance on dangerous bioweapons-related knowledge, so Anthropic was unable to rule out the need for ASL-3 safeguards while finding the model substantially below ASL-4. Claude Sonnet 4 remained below the ASL-3 thresholds of concern.

The single most informative evaluation is the bioweapons acquisition uplift trial. Groups of 8 to 10 participants had up to two days to draft a comprehensive bioweapons acquisition plan; the control group had only basic internet resources, the model-assisted group had Claude with safeguards removed. Deloitte graded the outputs. The threshold analysis said total uplift at or above 5 times would create significant additional risk and uplift at or below 2.8 times would keep risk acceptable. Control participants scored 25% plus or minus 13%; Claude Opus 4 participants scored 63% plus or minus 13%; Claude Sonnet 4 participants scored 42% plus or minus 11%. That is uplift of 2.53 times and 1.70 times respectively, against 1.53 times for Claude Sonnet 3.7. And, as the card says, all participants hit critical failures.

The other ASL-3 evaluations: Deloitte expert red teaming found substantially increased risk in certain parts of the acquisition pathway for Opus 4, while both models continued to make critical errors that would have prevented real-world success for many actors. Long-form virology tasks, developed with SecureBio, Deloitte and Signature Science, put Opus 4 at 0.83 and 0.720 on the two tasks against a rule-out threshold below 50% and a rule-in threshold above 80%. Multimodal virology is treated as saturated, because even Claude Sonnet 3.7 exceeded the 22.1% expert baseline. On 33 bioweapons

knowledge questions, both Claude 4 models beat experts on 17 of 33, below the 27 of 33 threshold. On DNA synthesis screening evasion, all models could either assemble viable plasmids or evade screening, but none could do both.

For ASL-4: creative biology produced comparable mean scores (0.45 Opus 4, 0.37 Sonnet 4, 0.41 Sonnet 3.7); 4 of 6 short-horizon computational biology tasks clearly scored below the rule-out bar for every model; and expert red teaming concluded Opus 4 is unlikely to uplift experts to a substantially concerning degree.

Why it's assigned

It is the Further-tier primary source behind 1.3 #5's headline claim, and the course wants her to see the actual numbers rather than the press summary. For the day's second question it does something none of the other readings do: it shows what measuring AI biological risk looks like in practice, including the thresholds, the graders, the human baselines and the places the measurement itself is admitted to be weak. Anyone who later argues about AI-bio policy needs to know that this evidence base is a handful of trials with wide error bars.

Words to know

- **CBRN** — chemical, biological, radiological and nuclear. The card's umbrella term, though it says it

concentrates on biology.

- **ASL (AI Safety Level)** — the tiered capability thresholds in Anthropic's Responsible Scaling Policy, each triggering particular safeguards. ASL-3 is about novices with basic technical backgrounds; ASL-4 is about state programmes.
- **Rule-in and rule-out thresholds** — the scores above which a capability is treated as present and below which it is treated as absent, with a middle zone requiring more evaluation. A result in the middle is what "unable to rule out" means.
- **Helpful-only model** — a version with harmlessness safeguards removed, used so that refusals do not hide the underlying capability. Most of these evaluations use one, which is why the results are not what a public user would get.
- **Elicitation** — the effort to get the best performance out of a model, through tools, agentic harnesses and iteratively refined prompting. The card is explicit that it elicits hard, so results are a ceiling rather than a typical case.
- **Agentic harness** — a software setup giving a model extra tools to complete multi-step tasks. Improvement here, rather than in knowledge, is what the card flags about Opus 4.
- **Uplift** — how much better someone does with model access than without. The trial's 2.53 times is a ratio of graded plan scores, not a measure of real-world success.

- **Information hazard** — knowledge that could cause harm simply by existing or being disclosed. It is why the creative biology evaluation uses harmless biological systems as a proxy, and why some thresholds and results are withheld.
- **Data leakage** — when a model can score well by recalling published material rather than doing the task. The card found this in some computational biology evaluations and discounted those results.

Numbers in this reading

- Bioweapons acquisition uplift trial: groups of 8 to 10 participants, up to two days; thresholds of total uplift at or above 5 times (significant additional risk) and at or below 2.8 times (acceptable), or raw uplift at or above 0.8.
- Trial scores: control 25% plus or minus 13%; Claude Opus 4 63% plus or minus 13%; Claude Sonnet 4 42% plus or minus 11%. Uplift 2.53 times and 1.70 times; Claude Sonnet 3.7 at 1.53 times on Deloitte's updated rubric. All participants hit critical failures.
- Long-form virology: rule-out below 50%, rule-in above 80%. Claude Opus 4 scored 0.83 ("max of 5") on task 1 and 0.720 ("pass@5") on task 2; Claude Sonnet 4 scored 0.55 and 0.635.
- Multimodal virology (SecureBio's VCT, multiple-select variant): expert baseline 22.1%; all models

tested have surpassed it, so the evaluation is treated as saturated.

- Bioweapons knowledge questions: 33 questions written by Deloitte staff; threshold of beating experts on more than 80%, that is 27 of 33. Claude Opus 4 and Claude Sonnet 4 each answered 17 of 33 better than experts; Claude Sonnet 3.7 and Claude Sonnet 3.5 (new) each scored 18 of 33.
- DNA synthesis screening evasion: 10 different pathogenic agents tested, six criteria of varying importance. No model completed the end-to-end task.
- LAB-Bench: four tasks (FigQA, ProtocolQA, SeqQA, CloningScenarios). Claude Opus 4 beat all other models on three and reached or exceeded the human baseline on those, while performing substantially below the human baseline on FigQA.
- Creative biology: mean 0.45 plus or minus 0.08 (Opus 4), 0.37 plus or minus 0.08 (Sonnet 4), 0.41 plus or minus 0.1 (Sonnet 3.7). Best-of-30 scores: 0.74 (Sonnet 4) and 0.81 (Opus 4). SecureBio's own researchers expect human experts to score approximately 10%.
- Short-horizon computational biology: 4 of 6 evaluations clearly below the rule-out bar for all models; for task 2, 50% of outputs across Opus 4 and Sonnet 4 performed above the lower bound.
- Red teaming durations: three days for ASL-3, two days for ASL-4 with one bioengineering and

biosecurity expert.

- Nuclear and radiological risk: no internal evaluations; a formal partnership with the US Department of Energy's National Nuclear Security Administration since February 2024, whose results are not published.

What the author is unsure about

Anthropic documents its own uncertainty in unusual detail. Some CBRN evaluation thresholds and results are sensitive and not disclosed publicly, though shared with external red-teaming and pre-deployment partners. ASL-4 threat-model work is described as still preliminary and subject to revision. The card says it remains uncertain about the relative importance of hands-on lab skills versus theoretical knowledge, that surveyed experts generally agree tacit knowledge matters while some suggest its importance as a barrier may be overstated, and that it is not sure how uplift measured on an evaluation translates into real-world uplift, with the best estimate relying on a probabilistic model. It calls text-based uplift trials substantially weaker proxies for real-world scenarios, which involve tacit knowledge, materials access and actor persistence, and says it adopted a similar threshold anyway to be conservative. The bioweapons knowledge evaluation is described as having low scoring reliability and being highly variable because of human grading. The creative biology evaluation has

no clear threshold, no human baseline yet, and high answer-to-answer variance; the card explicitly warns of a "frog-boiling effect" where gradual changes mask the crossing of important thresholds, and declines to use that evaluation as a strong ASL-4 rule-out. Some computational biology evaluations had data leakage. It does not run chemical evaluations internally, relying on the UK AI Security Institute and the US AI Safety Institute, and does not run nuclear or radiological evaluations internally at all. For Claude Sonnet 4's uplift trial it used an earlier, likely less capable snapshot.

For the discussion

- A point worth making: the famous "on the cusp" language comes down to a measured uplift of 2.53 times against an acceptable-risk ceiling of 2.8 and a significant-risk floor of 5. The safeguards were triggered by a result inside the acceptable band, because Anthropic treated proximity as enough.
- A tension worth raising: Anthropic says it is uncertain how evaluation uplift translates into real-world uplift and that text-based trials are weak proxies, and its own rule-out decisions rest on exactly those numbers. The honesty is admirable and it also means the policy is resting on an admittedly weak instrument.
- A question worth asking: on the one evaluation closest to a real-world chokepoint, DNA synthesis

screening evasion, no model could both assemble a viable design and evade screening. If that is the binding constraint, should the course treat synthesis screening as the thing to defend above all else?

1.3-7 · CORE · COURSE TIME 25 MIN

The New Bioweapons: How Synthetic Biology Could Destabilize the World

Roger Brent, T. Greg McKelvey, Jr., and Jason Matheny · 2024

[Read the original paper →](#) · the course assigns: whole article (Foreign Affairs, September/October 2024)

Three authors from Fred Hutch and RAND argue that Cold War deterrence cannot work for biological weapons, so the only route is deterrence by defence: harden society until a biological attack is not worth conducting, the way cities were hardened against fire.

Average summary

The article opens with a metaphor that carries the whole argument. In cybersecurity, a penetration test is a simulated attack using an adversary's own tools; banks hire experts to break in and then report how to improve security. At the end of the last decade and the start of this one, human society itself was subject

to a kind of penetration test: **COVID-19**. The virus, an unthinking adversary, probed the world's ability to defend against new pathogens, and by the end of the test it was clear humanity had failed.

Their evidence for failure is reach, not death toll. COVID-19 went everywhere, from remote Antarctic research stations to isolated Amazonian tribes, through nursing homes and aircraft carriers, levelling frontline workers and heads of state alike. By the end of 2022, three of every four Americans had been infected at least once; in the six weeks after China ended zero COVID in December, over one billion of its people were infected. The primary reason the death toll was relatively modest, they argue, was not that society controlled the disease but that infection proved only modestly lethal. COVID-19 mostly burned itself out.

The history section runs from German sabotage of Allied transport animals in the First World War, through Unit 731 in Japanese-occupied Manchuria where thousands of prisoners were infected and killed, to the US Department of Defense's Project 112 in the 1960s, which dispersed spores in New York City subway tunnels and bacteria from boats in San Francisco Bay. Nixon halted the US programme in 1969 and Joshua Lederberg testified to Congress that biological weapons could become as deadly as nuclear ones but easier to construct, warning that

germ power would work the opposite way from nuclear nonproliferation. The Soviets signed the Biological Weapons Convention and defied it: 68 people died at Sverdlovsk in 1979, and the programme employed 60,000 people at its height. Then the non-state record: the Rajneesh movement's 1984 salmonella contamination, Aum Shinrikyo's 1995 sarin attack, and the 2001 anthrax letters that killed five.

The second half argues why nuclear-style deterrence fails: mutually assured destruction relies on widespread fear that biological warfare does not generate, on being able to identify the attacker, and on an adversary who wants to survive. None of the three holds. So the path to deterrence is a defence strong enough that attacks are not worth conducting, and the template is the end of major urban fires, which came less from fire brigades than from less combustible materials, engineering standards, building codes and liability insurance. Their applied version: develop vaccines and antivirals in days rather than months, with the capacity to vaccinate eight billion people within 100 days; stockpile and redesign personal protective equipment; install far-UV light purification and particulate filters in essential-services buildings; and govern the technology, with know-your-customer rules for suppliers of nucleic acids, strains and reagents.

Why it's assigned

It is the Core reading that turns the day from diagnosis to design, and it is where the course's four-layer story comes from. Brent, McKelvey and Matheny give Unit 1 its organising analogy (defence as building codes rather than firefighting), its core strategic claim (deterrence by denial, since retaliation cannot work for biology), and a preview of every later unit: governance and know-your-customer rules in Unit 2, warning systems in Unit 3, PPE and far-UV in Unit 4, and 100-day vaccines in Unit 5. It also pairs with 1.4 #1, which develops the same fire analogy at greater length.

In depth

COVID-19 as a penetration test

The opening analogy is precise. A penetration test is a simulated attack on a computer system's defences using an adversary's tools and techniques; banks regularly hire experts to break in and transfer money to unauthorised accounts, often by phishing login credentials from employees, and then present their findings and recommendations.

Human society got such a test in COVID-19, from an unthinking adversary, and **humanity had failed**. The authors list where the virus got: remote Antarctic research stations, isolated Amazonian tribes, nursing homes, aircraft carriers, frontline workers and heads

of state alike. The draconian lockdowns imposed by autocracies and the miraculous vaccines developed by democracies slowed but did not halt its spread. By the end of 2022, three of every four Americans had been infected at least once. In the six weeks after China ended its zero COVID restrictions in December, over one billion of the country's people were infected.

Then the sentence that reframes everything: the primary reason for the pandemic's relatively modest death toll was not that society had controlled the disease, but that viral infection proved to be only modestly lethal. In the end, COVID-19 mostly burned itself out.

That failure is sobering because the world faces a growing number of biological threats. Some, such as avian flu, come from nature. Plenty come from scientific advances: over the past 60 years researchers have learned to create viruses that evade immunity, to evolve existing viruses to spread more easily through the air, and to engineer viruses to make them more deadly. The authors state plainly that it remains unclear whether COVID-19 arose from such activities or entered the human population via interaction with wildlife, and say that either way, biological technology now boosted by artificial intelligence has made it simpler than ever to produce diseases.

Risky business: progress as a double-edged sword

Then the other edge. During the First World War, German military intelligence operatives used pathogens to attack the animals the Allies used for transport, sickening horses and mules in France and Romania, attempting to infect reindeer used by the Sami in Norway, and infecting corrals and stables in the United States full of animals bound for Europe. By the Second World War these had matured into weapons designed to kill humans. In Japanese-occupied Manchuria, the Japanese military officer Shiro Ishii presided over Unit 731, where forces tested biological weapons on humans, infecting and killing thousands of prisoners with anthrax, typhoid, paratyphoid, glanders, dysentery and bubonic plague. In the final days of the war Ishii proposed a full biological-warfare operation, titled Cherry Blossoms at Night, in which Japanese seaplanes would disperse bubonic-plague-infected fleas over major American West Coast cities. The chief of the army general staff vetoed it with a line the authors quote: if bacteriological warfare is conducted, it will grow from the dimension of war between Japan and America to an endless battle of humanity against bacteria.

Other countries kept going. In the 1960s the US Department of Defense launched Project 112, which experimented with mass distribution of offensive pathogens: the army dispersed spores in the tunnels

of the New York City subway and bacteria in aerosols from boats in San Francisco Bay, and sprayed chemicals from army planes over thousands of square miles, from the Rockies to the Atlantic and from Canada to the Gulf of Mexico. The logic was insurance against a Soviet nuclear attack: if Moscow neutralised Washington's nuclear arsenal, the United States could still devastate the Soviet Union with pathogens. By mid-decade the department had committed to developing lethal and incapacitating biological weapons, and by the end of the 1960s government scientists were producing sizable quantities of deadly bacteria and toxins devised, in the words of the microbiologist Riley Housewright, to confound diagnosis and frustrate treatment.

Civilian researchers were terrified by this and pushed back, finding a receptive audience in the White House. In 1969 President Richard Nixon halted the US biological weapons programme and called for an international treaty. Shortly after, Joshua Lederberg, a Nobel Prize-winning biologist, testified before Congress in support of a global ban: biological weapons could become just as deadly as nuclear ones, but easier to construct. His quoted warning is the article's sharpest historical point. Nuclear weaponry "has been monopolized by the great powers long enough to sustain a de facto balance of deterrence and build a security system based on

nonproliferation", Lederberg said. "Germ power will work just the other way."

Washington's main adversary was not persuaded. In 1971, as the world haggled over a treaty, the Soviet Union released a weaponized strain of Variola major, the smallpox virus, on an island in the Aral Sea, causing an outbreak in present-day Kazakhstan. It was contained by heroic efforts from Soviet public health officials, but only because the region was sparsely populated and most Soviet citizens had been vaccinated. Later that year the Soviet Union and the United States agreed to the Biological Weapons Convention; the UN General Assembly universally commended it, and in 1972 it opened for signing in London, Moscow and Washington. The Soviets defied it: in 1979, 68 people died in Sverdlovsk, present-day Yekaterinburg, after spores from a clandestine anthrax project were released. Moscow had no other clear-cut accidents, but maintained a programme until the country fell apart, employing 60,000 people at its height according to defectors. In 1991, US and British representatives visited some facilities and saw rows of vessels and bioreactors capable of producing thousands of litres of high-titer smallpox, which could be pumped through refrigerated pipes into bomblets and loaded onto missiles.

The Biological Weapons Convention had a second problem: it did not constrain private groups and

individuals, and the authors run through the Rajneesh movement's 1984 salmonella contamination, Aum Shinrikyo's 1995 sarin attack after failed anthrax attempts, and the 2001 anthrax letters that killed five. Then they make the move that defines the piece. The relatively small scale of these incidents could be taken as evidence that terrorists and states are too constrained, by technical difficulties or existing laws, to inflict mass biological damage. They call that perspective too optimistic. Instead, the incidents show that current international agreements and public health measures cannot prevent such attacks, and demonstrate that it is wrong to assume states and terrorists lack the will or the means to build biological weapons. Some individuals and groups do face barriers, such as an inability to access the right labs or facilities, but thanks to relentless technological advances those barriers are falling apart.

For better and for worse: the two technology revolutions

In 2012 a group of scientists led by Emmanuelle Charpentier and Jennifer Doudna published an article in *Science* describing CRISPR-Cas9, an engineering system that uses human-made chimeric RNA to edit genetic material. It joined an already formidable toolbox: classical recombinant DNA (invented in the 1970s), the polymerase chain reaction or PCR (1980s), and synthetic DNA (also in use from the 1980s). Together these have created an explosion of human

ingenuity powering discovery and medicine. In December 2023 the FDA approved a complex CRISPR-based gene therapy as a cure for sickle cell anemia, a devastating illness afflicting millions.

The second revolution is AI. Large language models such as those from ChatGPT and Claude grow more sophisticated with each iteration and are used daily by thousands of lab workers to accelerate their work, in part by providing a wealth of useful guidance on technical questions. In 2020 AI researchers created AlphaFold, which effectively solved a Holy Grail problem in biology: predicting the three-dimensional structure of a protein from its amino acid sequence.

For would-be bioterrorists, the authors argue, these systems could ease the path to mayhem. The largest AI models appear to have been trained on the entirety of the life sciences' published knowledge; most of that was already on the internet, but no human could consume, process and synthesise all of it. Present systems can also design new proteins, which enables the design of dangerous pathogens, and execute laboratory operations. Some computer scientists are working on automated systems that can carry out laboratory tasks, and if those succeed, a malevolent actor could create a deadly new pathogen simply by hijacking such automated facilities.

Their asymmetry argument is the one to remember: generating pathogens is cheaper than defending

against them. The capital costs of the facilities and materials needed to make a new disease are low, while responding to an epidemic involves expansive testing and detection networks, vast quantities of personal protective equipment, socially disruptive lockdowns, and an apparatus to develop, manufacture and distribute treatments and vaccines. If bad actors do release a viral pathogen, it could infect vast swaths of the population in far less time than officials would need to detect and identify the threat and start fighting back.

They close the section on state intent. The thought of spending billions trying to stop another pandemic should be enough to deter states from weaponizing biology, yet in April 2024 the US State Department assessed that North Korea and Russia have offensive biological weapons programmes and that China and Iran are pursuing biological activities that could be weaponized. All are parties to the Biological Weapons Convention.

Deterrence by defense

Here is the strategic heart. During the Cold War, nuclear powers avoided catastrophe largely thanks to mutually assured destruction: politicians recognised that a single nuclear attack might trigger a planet-ending retaliation, or as Reagan and Gorbachev declared in 1985, a nuclear war cannot be won and must never be fought.

First, mutually assured destruction relies on fear, which was widespread in the nuclear era and is not as prevalent toward biological warfare. The current threat depends on continued breakneck technological progress and on inventions without precedent, which makes the risks hard to fully grasp. And unlike Hiroshima and Nagasaki, no biological attack has been a world-historical event that attracts enduring attention.

Second, it depends on being able to identify the attacker. With nuclear weapons that is easy enough. States could release biological weapons and evade detection, and therefore retaliation; a government could secretly release a dangerous virus and blame it on any number of other states, or on non-state actors.

Third, non-state actors really could release deadly pathogens. No government wants to risk the annihilation of its country, but plenty of terrorists care little about survival, and they now have access to the materials, equipment, knowledge and technical capability needed. The authors return to Lederberg, who warned in 1969 that the consequences of unchecked biological proliferation would be akin to making hydrogen bombs available at the supermarket. Their gloss: the world of 2024 is full of supermarkets, well stocked with bomb-making materials.

So the path to deterrence is not retaliation but a defence so strong it makes biological attacks not

worth conducting.

The fire template

Their historical template is the end of major urban fires. For most of recorded history the cities of the world were periodically consumed by massive conflagrations that razed their cores, and in the nineteenth century the frequency fell dramatically. Part of that was better response systems: professional firefighting forces and fire hydrants. But mostly the reduction was driven by mundane steps: less combustible building materials, engineering standards and building codes, and requirements for liability insurance, which discouraged risky behaviour. When states created sharper definitions of negligence, making it easier to launch civil suits for accidental fires, people became even more cautious.

Translated: governments built fire departments and hydrants, so now they need systems that can rapidly develop vaccines, antiviral drugs and other medical interventions. But rapid response will not be enough. The world could, and must, develop the ability to vaccinate its eight billion people within 100 days of an outbreak, faster than it took the United States to fully vaccinate 100 million people against COVID-19. And even that would not suffice for a pathogen spreading at the pace of the coronavirus's Omicron variant.

So policymakers must also take steps akin to better building codes, which make it harder for pathogens to spread. Bigger stockpiles of personal protective equipment, since masks, gloves and respirators are key to stopping transmission, with preparatory contracts signed in advance. Subsidies for industrial bases so they can surge production. Instructions to manufacturers to redesign PPE to be cheaper, more effective and more comfortable. Priority access for people who work in essential services. And microbicidal far-UV-light purification systems and particulate filters in those sectors' buildings. Combined, they say, these measures would substantially reduce the risk that outbreaks grow into societally destabilising events.

Step by step: governing the technology

The final layer goes beyond responses and defences to governing new technologies, and the authors say it may ultimately be the only way to actually prevent a mass biological attack.

Their tools: deny funding to or outright ban particular experiments; require licences for people and facilities before certain kinds of work; and oversee future lab automation more thoroughly. Beyond that, shape the ecosystem. Require firms selling nucleic acids, strains, reagents and other life-sciences equipment to adopt know-your-customer rules, confirming customers' identities and the nature of their activities

and ensuring goods ship only to known, legitimate sites. They point out the precedent: many governments have long forced financial institutions to follow know-your-customer rules to keep money out of criminal networks. Policymakers should also better regulate conduct, devising new ways to detect prohibited biological activity so law enforcement and intelligence agencies can head off attacks before they happen.

And biodefense policy must be crafted with AI in mind, starting today. Before releasing large language models, companies invent and install safeguards such as redlines users cannot cross. Their worked example of how thin those can be: ChatGPT-4 and Claude 3.5 Sonnet refuse to answer direct questions on how to evolve a virus to kill farm animals, but if users ask for technical guidance on how to engage in such directed evolution without using the word "kill", these models will give good guidance. AI models therefore need additional safeguards against handing out dangerous information, and governments should help create them.

The closing note is deliberately not despairing. The world has avoided existential catastrophe before. The Cold War may not provide a template for today's challenges, but its history is still proof that society can contain dangerous inventions: then, as now, the world faced an innovation developed by human

ingenuity that imperiled civilization; then, as now, states could not eliminate it; and governments succeeded in preventing the worst through concepts and systems that kept the risk to a minimum. They give the last word to John von Neumann, the mathematician and physicist who helped guide US nuclear policy: "For progress, there is no cure. Any attempt to find automatically safe channels for the present explosive variety of progress must lead to frustration. The only safety possible is relative, and it lies in an intelligent exercise of day-to-day judgment." And: "To ask in advance for a complete recipe would be unreasonable. We can specify only the human qualities required: patience, flexibility, intelligence."

Words to know

- **Penetration test** — a simulated attack using an adversary's own methods, run to find weaknesses before a real attacker does. The authors use it to argue COVID-19 told us what our defences are worth.
- **Deterrence by defense (or denial)** — making an attack pointless by being too hard to hurt, rather than by threatening retaliation. This is the article's central strategic proposal and the reason the fire analogy fits.
- **Mutually assured destruction** — the Cold War logic that any nuclear attack would trigger

annihilating retaliation. The article's job is to explain why its three preconditions fail for biology.

- **CRISPR-Cas9** — a gene-editing system described in 2012 that uses human-made chimeric RNA to edit genetic material. Named here with recombinant DNA, PCR and synthetic DNA as the toolbox that changed what is possible.
- **Know-your-customer rules** — the obligation on a supplier to confirm who a customer is and what they are doing. Long imposed on banks, and proposed here for sellers of nucleic acids, strains and reagents.
- **Redlines (in AI models)** — safeguard boundaries users are not supposed to cross. The article's example of asking the same question without the word "kill" shows how a redline can be worded around.
- **Far-UV light purification** — germicidal ultraviolet light used to clean indoor air. Proposed as part of the biological equivalent of a building code.
- **High-titer** — at high concentration of virus. Used to describe what Soviet bioreactors could produce, and it is the detail that makes the scale real.

Numbers in this reading

- By the end of 2022, three of every four Americans had been infected with SARS-CoV-2 at least once.
- In the six weeks after China ended zero COVID in December, over one billion people in the country

were infected.

- The Black Death killed one of every three people in Europe; a worst-case synthetic pathogen might exceed that toll.
- Over the past 60 years, researchers developed sophisticated understandings of molecular and human biology.
- Sverdlovsk, 1979: 68 people died. The Soviet programme employed 60,000 people at its height, with bioreactors capable of producing thousands of litres of high-titer smallpox.
- The 1971 Aral Sea release of weaponized Variola major caused an outbreak in present-day Kazakhstan.
- 1984 Rajneesh salmonella contamination: no deaths, hundreds ill. 1995 Aum Shinrikyo sarin attack: thousands injured. 2001 anthrax letters: five killed.
- CRISPR-Cas9 published in 2012; FDA approved a CRISPR-based sickle cell cure in December 2023 at \$2.2 million per person.
- Horsepox was re-created in 2018 by one individual on a three-person team.
- AlphaFold created in 2020.
- April 2024 US State Department assessment: North Korea and Russia have offensive biological weapons programmes; China and Iran are pursuing activities that could be weaponized.

- The target the authors set: vaccinate eight billion people within 100 days of an outbreak, faster than the United States took to fully vaccinate 100 million people against COVID-19.
- Reagan and Gorbachev's declaration, 1985. Lederberg's congressional testimony and Nixon's halt, 1969.

What the author is unsure about

They are explicit on the origin question: it remains unclear whether COVID-19 arose from laboratory activities or entered the human population via interaction with wildlife. The catastrophic scenarios are framed as possibilities ("might exceed", "could be catastrophic"), not predictions. The claim about AI easing the path to mayhem is hedged with "could", and they concede the underlying knowledge was already on the internet, so the uplift is in synthesis rather than access. They acknowledge their own prescriptions carry costs: targeting suppliers could slow valuable biomedical research and development, and some governance measures risk slowing legitimate research, so policymakers must be thoughtful. They admit that even a 100-day global vaccination capability would not suffice against a pathogen spreading at Omicron's pace. And rather than offering certainty they close on von Neumann's claim that only relative safety is possible, which is a statement that there is no complete recipe.

For the discussion

- A point worth making: their best line is that COVID-19's toll was modest because the virus was only modestly lethal, not because we controlled it. Everything in the course follows from accepting that one sentence.
- A tension worth raising: the fire analogy cuts in a direction the authors mostly leave implicit. Cities were made safe largely by liability, insurance and sharper definitions of negligence, not by technology. The biological equivalent would mean someone being legally and financially answerable for a release, and nobody in this field currently is.
- A question worth asking: they propose know-your-customer rules for suppliers of nucleic acids, strains and reagents on the banking model. Who in biology plays the role the banks play, and what happens to the suppliers who are not in a jurisdiction that imposes it?

1.3-8 • CORE • COURSE TIME 5 MIN

Securing Civilisation Against Catastrophic Pandemics

Anjali Gopal, William Bradshaw, Vaishnav Sunil and Kevin M. Esvelt • 2023

[Read the original paper →](#) • the course assigns: 'printed pages 10 to 12, section II "Two scenarios for catastrophic pandemics" (PDF pages 11 to 13)'

Two different kinds of pandemic could bring down a society: a "Wildfire" agent so lethal and contagious that essential workers stop showing up, and a "Stealth" agent so mild at first that it infects most of humankind before anyone knows it exists.

Average summary

The course assigns three pages of this Geneva Paper, and they are the three that the rest of the week refers back to. The authors describe two distinct mechanisms by which a pandemic pathogen could cause societal collapse, and the collapse in both cases runs through the same bottleneck: **essential workers**.

Scenario 1, "Wildfire". One or more sufficiently debilitating and transmissible pandemic agents trigger civilisational collapse by interrupting the distribution of food, water, power and law enforcement. COVID-19 showed that some high- and middle-income nations can contain an agent temporarily by locking down, but a lockdown only holds if essential goods and services keep moving, and those are provided by workers who were infected at particularly high rates during COVID-19. The vast majority of them have no protection beyond N95 masks, which theoretically block 95% of particles down to 0.3 microns but in practice have been found to block only 57 to 86% of coronavirus particles worn

loosely, and 79 to 90% worn correctly. If a nation loses too many essential workers, to death, debilitation or refusal to work through fear, households will not get the food, water, heating and protection they need, and law and order break down.

To qualify as a Wildfire agent, a highly lethal overt pathogen must be transmissible enough to infect most essential workers in high-income countries while the rest of the population locks down; the authors note comparable pathogens exist in other species and could be engineered in humans. Their most unsettling point is about perception: a visibly severe Wildfire scenario plus misinformation could trigger collapse even if the actual risk of infection is low, because workers who believe their peril is unacceptable will reasonably decline to risk carrying a deadly virus home to their families, even if cases are falling. Perception, at least among essential workers, is just as important as epidemiological reality.

Scenario 2, "Stealth". An initially mild or asymptomatic agent with a long incubation period infects most of humankind before its harmful effects become apparent. SARS-CoV-2, tuberculosis and HIV killed more people in 2021 than any other pathogens, and all three show some minimally symptomatic transmission or latency, but only SARS-CoV-2 spreads rapidly and only HIV natively has both a long incubation period and a very high infection fatality

rate. A pathogen combining these traits, colloquially a "fast-spreading HIV-equivalent", could infect much of humankind before anyone noticed, then debilitate or kill most essential workers. Today it would almost certainly go undetected, because patients with common colds are rarely sequenced, so nobody would have cause to look for signs of adversarial engineering in the genome.

Their hardest question is political: if no one is in hospital, no one has died, and only the scientific community is alarmed, will governments take politically costly actions? And if people later understand they are infected, and so is almost everyone around them, how they would react is unclear. The authors say preserving hope in the form of newly devised treatments could be critical to maintaining order.

Why it's assigned

These three pages are the course's worst-case vocabulary. Units 3 and 4 refer back to them directly: the Stealth scenario is why detection has to be pathogen-agnostic rather than looking for known threats, and the Wildfire scenario is why Unit 4 is about blocking transmission and protecting the people who have to keep working rather than only about vaccines. Read in Unit 1, they also answer the day's third question from the other end. Instead of

describing a pandemic-proof world, they describe precisely what it has to survive.

In depth

Scenario 1: the "Wildfire" pandemic

The authors' one-line definition: one or more sufficiently debilitating and transmissible pandemic agents trigger civilisational collapse by interrupting the distribution of **food, water, power and law enforcement**.

The mechanism starts from what COVID-19 demonstrated, which is that some high- and middle-income nations may have the resources to temporarily contain pandemic agents by locking down their populations to limit human contact. But maintaining a lockdown requires the continued distribution of essential goods and services, and those are generated and provided by essential workers, many of whom were infected at particularly high rates during COVID-19.

The protection those workers have is the weak point, and the authors give the numbers rather than an adjective. The vast majority of essential workers lack protective equipment beyond N95 masks, which theoretically block 95% of particles down to 0.3 micrometres, but in practice have been found to block only 57 to 86% of coronavirus particles when worn loosely and 79 to 90% when worn correctly.

The cascade from there is stepwise. If a nation loses too many essential workers, whether to death, debilitation or refusal to work through fear of contagion, some households will fail to receive the food, water, heating and protection they need to survive. People forced to leave their homes in search of vital supplies will spread the virus and further disrupt essential services. As law and order increasingly break down, societies will collapse.

The threshold a Wildfire agent has to meet

To qualify, a highly lethal overt pathogen must be transmissible enough to infect most essential workers in high-income countries while the rest of the population locks down. If too many essential workers die or refuse to work, society will collapse. The authors add one line that does a lot of work: comparable pathogens exist in other species and could be engineered in humans. The scenario is not hypothetical biology; it is existing biology in the wrong host.

The part about perception

This is the passage most worth remembering, because it breaks the usual assumption that defence is a technical problem.

Crucially, the authors write, a visibly severe Wildfire scenario and the spread of misinformation could trigger societal collapse even if the actual risk of

infection is low. If essential workers perceive their own level of peril to be unacceptably high for any reason, they will reasonably decline to risk contracting an exceptionally deadly virus and carrying it home to their families. This may occur even if case numbers are falling among the population as a whole. Their conclusion: perception, at least among essential workers, is just as important as epidemiological reality.

Note the word "reasonably". The authors are not describing panic or irrationality. They are describing a correct individual decision that adds up to collapse.

Figure 3

The section's illustration is described as an artistic depiction of the consequences of a Wildfire pandemic, credited to Pieter Bruegel the Elder (around 1525 to 9 September 1569). It is there to make the societal-breakdown end state imaginable rather than to carry data.

Scenario 2: the "Stealth" pandemic

The definition: an initially mild or asymptomatic agent with a long incubation period infects most of humankind before its harmful effects become apparent.

The authors build the scenario out of real pathogens rather than invention. SARS-CoV-2, tuberculosis and

HIV killed more people in 2021 than any other pathogens. All three exhibit some form of minimally symptomatic transmission or latency. But only SARS-CoV-2 spreads rapidly, and only HIV natively exhibits both a long incubation period and a very high infection fatality rate.

A lethal pathogen that combines these traits, colloquially a "fast-spreading HIV-equivalent", could infect much of humankind before anyone noticed its existence, then debilitate or kill most essential workers, triggering societal collapse.

Then the detection failure, which is specific and is the reason Unit 3 exists. Today, such an event would almost certainly go undetected, because patients with common colds are rarely sequenced, so no one would have cause to look for signs of adversarial engineering in the genome sequence of the pathogen. The gap is not that we lack the technology to see it; it is that nobody is looking at the samples where it would be visible.

They add a second obstacle that holds even if detection worked. SARS-CoV-2 also demonstrated that it is exceptionally difficult to contain a pandemic pathogen when a substantial fraction of the population does not believe that it poses a threat.

The figure: timeline of a Stealth pandemic

Figure 4 is the piece of this section that the rest of the course quotes most. It is a timeline with three marks: **4 months**, **3 years** and **6 years**. The first stretch is labelled mild or no symptoms; the later stretch is labelled delayed-onset death or disability.

What it shows: a Stealth agent with transmissibility comparable to SARS-CoV-2 would rapidly spread across the world. With mild or non-existent symptoms there would be no attempts to limit its spread until years later, when the first of those infected begin to suffer debilitating illness or death. What to notice is the shape of the problem rather than any single number: the entire window in which action is cheap falls inside the stretch where there is no visible evidence of harm. The caption states the consequence plainly: if too many people have already been infected and effective treatments cannot be developed in time, civilisation will collapse.

The political problem

The authors pose it as a question, in their own words: if no one is in hospital, no one has died, and only the scientific community is alarmed, will governments take politically costly actions to stop the transmission of a pandemic agent that might or might not be harmful?

Note the "might or might not". They are not assuming the warning is certain. The scenario they are

describing asks politicians to impose visible costs now on the strength of a prediction that will only be confirmed years later.

With adequate preparation, they say, much of the population might be persuaded that infection would entail delayed disability or death. But even then, those people may still struggle to protect themselves from being infected by people who do not believe the pandemic is real. The defence is partly social, and it can be defeated by the fraction of society that opts out.

What happens when it becomes visible

The last part of the section is unusual for a security paper, because it is about despair.

As more and more victims began to suffer the effects, people would come to understand that they had become infected and would likely suffer the consequences, as would almost everyone around them. How they would react, the authors write, is unclear. Perhaps they would come together to do what they can to preserve humankind's future, or they might withdraw in despair.

Their one intervention here is not technological: preserving hope in the form of newly devised treatments could be critical to maintaining order. That is the thread the report's later Defences section picks up, where the recommendation is to offer fast grants

for every promising therapeutic approach partly so that there is credible hope to communicate.

Either way, they conclude, if most essential workers become infected and the scientific community cannot devise a treatment before they suffer debility or death, societies will collapse.

Words to know

- **Wildfire agent** — in this report's terms, a highly lethal and overt pathogen transmissible enough to infect most essential workers in a high-income country even while everyone else locks down. The report's later sections define it quantitatively, but on these pages it is defined by the effect.
- **Stealth agent** — an agent that spreads with mild or no symptoms and causes severe harm years later; the report's shorthand is a fast-spreading HIV-equivalent.
- **Essential workers** — those who must deliver food, water, power and law enforcement without interruption. Both scenarios collapse through this group, which is why the term is load-bearing rather than descriptive.
- **Incubation period** — the time between infection and the appearance of symptoms. A long one is what makes the Stealth scenario undetectable while it is still containable.

- **Latency** — a pathogen persisting in the body without causing illness, sometimes for decades. Tuberculosis and herpesviruses are the natural examples that make the Stealth concept plausible.
- **Infection fatality rate** — the share of all infected people who die, including those never diagnosed. HIV's high rate combined with its long incubation is the half of the Stealth profile nature already supplies.
- **Overt (versus stealth) pathogen** — one whose harm is immediately visible. The distinction matters because every current defence assumes the harm announces itself.
- **Adversarial engineering** — deliberate modification intended to cause harm, which can sometimes be read in a genome sequence. The report's point is that nobody is sequencing the samples where it would show up.

Numbers in this reading

- N95 masks theoretically block 95% of particles down to 0.3 micrometres.
- In practice they have been found to block only 57 to 86% of coronavirus particles when worn loosely, and 79 to 90% when worn correctly.
- The four services that define primary collapse: food, water, power and law enforcement.
- SARS-CoV-2, tuberculosis and HIV killed more people in 2021 than any other pathogens.

- Stealth timeline (Figure 4): 4 months, 3 years, 6 years, with mild or no symptoms giving way to delayed-onset death or disability.

What the author is unsure about

Within these three pages the authors mark three open questions and one deliberate hedge. They say it is unclear how people would react once a Stealth agent's harms became apparent, offering two possibilities (coming together, or withdrawing in despair) without choosing. They pose the question of whether governments would act on a scientific warning alone as a question, not an assertion, and frame the pathogen as one that "might or might not be harmful". They say much of the population "might" be persuaded with adequate preparation, and immediately add that persuaded people may still be unable to protect themselves from those who are not. And their claim that comparable pathogens could be engineered in humans is a statement about what is possible, not about what exists. What they state flatly, without hedging, is the conditional: if too many essential workers die or refuse to work, societies will collapse.

For the discussion

- A point worth making: the collapse in both scenarios does not run through hospitals, it runs through supply chains. The variable that decides

whether a society survives is whether the people who move food, water and power believe they can do it without killing their families.

- A tension worth raising: the authors say perception is just as important as epidemiological reality, and they call the workers' decision reasonable. That means the standard public health move of reassuring people would be exactly the wrong intervention, because what is needed is protection good enough to deserve the reassurance.

1.3-9 · CORE · COURSE TIME 15 MIN

Mirror Life: Brief of the Scientific Advisory Board

UN Scientific Advisory Board · 2026

[Read the original paper →](#) · the course assigns: whole brief (6 pages)

The UN Secretary-General's Scientific Advisory Board draws a hard line between mirror molecules, which are already useful and low-risk, and self-replicating mirror life, which does not exist, is probably a decade away, has almost no known benefit, and could be catastrophic.

Average summary

The brief opens with six numbered summary points that are the whole argument in miniature. All living things are made of molecules that come in two mirror-image forms, like left and right hands; this is **chirality**. Mirror biology creates biological molecules with the opposite handedness, for example proteins built from right-handed rather than left-handed amino acids. There are real practical benefits, mainly the stability and durability of mirror molecules against degradation, and mirror DNA, RNA and some proteins are already produced for medical and biotechnical purposes. Mirror life, by contrast, is the still hypothetical idea of making self-replicating organisms out of mirror molecules; it has not been created and is likely a decade away. While the risks of mirror molecules are low, the risks of self-replicating mirror life are extraordinarily high. And given how fast the science is moving, a dedicated global forum to discuss risks and develop governance frameworks would be advisable.

The biology is explained simply. Proteins are built from left-handed amino acids while DNA and RNA are built from right-handed sugars, and that orientation lets them interact, fitting together like keys in locks. Mirror molecules have identical physical and chemical properties but are incompatible with natural life in important ways: our enzymes cannot process them, our immune sensors are less able to detect them, natural organisms cannot metabolise them efficiently,

and natural viruses cannot infect them. That incompatibility is what makes them useful as medicines and diagnostics, because they last far longer in blood, tissues and environmental samples. One application is mirror therapeutics, including short RNA strands known as Spiegelmers, several of which have entered preclinical or clinical development.

The risk case turns on self-replication. A mirror organism could use basic environmental resources that have no handedness, like oxygen or simple carbon-based substances, while largely avoiding attack by predators, immune systems or rival microbes. So it could bypass many of our most important biological defences, resist known antibiotics, infect plants and animals and colonise environments, behaving like a highly resilient invasive species with few built-in limitations on its growth. It would grow more slowly than its natural counterpart, because fewer nutrients would be viable, but could still theoretically spread across the planet. And it would be very hard to see: almost all diagnostic tests are designed to find regular life forms, so they probably would not detect mirror DNA or proteins.

The considerations for the UN system are the policy core: whether current international instruments such as the Cartagena Protocol on Biosafety and the Biological Weapons Convention are sufficient; whether monitoring and preparedness are adequate,

given that mirror life may fall outside current bio-surveillance; what guardrails could protect research on mirror molecules while ensuring mirror life is not created, including red lines such as preventing the creation of a mirror ribosome; and how to encourage global multi-sectoral collaboration.

Why it's assigned

It is the final Core reading of the day and the clearest single case study for the unit's second question.

Mirror life is the point where "nature sets a baseline, not a ceiling" stops being an argument and becomes a specific, named, decade-away capability. It also introduces the governance problem the course returns to in Unit 2 and Unit 6: there is widespread scientific agreement that mirror life should not be made, and as of today there is no global process to turn that agreement into governance. For the day's third question it supplies the hardest version of a pandemic-proof world, because a mirror pathogen would defeat immunity, antibiotics and diagnostics all at once.

In depth

The six summary points

The brief front-loads its conclusions, and they are worth having in order. One: all living things are made of molecules, many of which come in two mirror-image forms, like left and right hands; this is called

chirality, and although the two forms contain the same components, their mirrored shape can make them behave differently. Two: mirror biology refers to the study and practice of creating biological molecules that have the opposite handedness of their naturally occurring counterparts. Three: there are many practical benefits to mirror biology, including stability and durability against degradation in the natural world, and mirror DNA, RNA and some proteins are currently produced for medical and biotechnical purposes. Four: mirror life refers to the still hypothetical idea of creating self-replicating organisms, like a bacterium, out of mirror biological molecules; mirror life has not been created and is likely a decade away. Five: while the risks of mirror molecules are low, the risks associated with self-replicating mirror life are **extraordinarily high**, since mirror cells could spread through our environment and human populations, resisting important natural barriers and predators, and in some scenarios could pose a catastrophic threat to humanity. Six: given the rapidly advancing science, a dedicated global forum to discuss risks and develop governance frameworks to prevent the risks associated with mirror life would be advisable.

The brief carries a disclaimer on every page: the views expressed do not necessarily reflect those of the United Nations, its affiliated agencies and programmes, or the scientific community writ large,

but reflect the independent assessment of the Scientific Advisory Board.

From mirror biology to mirror life

All natural living things, from bacteria and plants to animals and people, are built from building blocks that can exist in two mirror-image forms, like left and right hands. The property is called chirality, meaning handedness. Proteins are built from left-handed amino acids, while DNA and RNA are built from right-handed sugar molecules, and the orientation of these molecules is what allows them to interact, fitting together like keys in locks.

Mirror biology creates versions of these molecules flipped to their mirror image. A mirror protein is composed of right-handed amino acids; mirror DNA uses left-handed sugars.

The crucial property is incompatibility rather than difference. Mirror molecules have identical physical and chemical properties to natural molecules but are incompatible with natural life in important ways: the natural enzymes in our bodies cannot process them, and our immune sensors are less able to detect them. Natural organisms would be unable to metabolise mirror cells efficiently, and natural viruses would not be able to infect them.

Figure 1 is described as natural versus mirror-image building blocks: natural DNA and proteins use one

handedness, while mirror biology flips them to the opposite form. It is there to establish the single idea the whole brief rests on.

What mirror molecules are already good for

Non-reproducing mirror molecules have been developed for use in biotechnology, medicine and materials, and their stability and resistance to natural degradation is the key feature. Mirror DNA and RNA cannot be broken down by ordinary enzymes, and mirror proteins similarly resist digestion and breakdown, so they can last longer in the human body than natural molecules that break down more quickly.

The flagship application is mirror therapeutics, including short RNA strands known as Spiegelmers. These strands accurately bind disease-related proteins while avoiding immune detection and enzymatic degradation, and several have entered preclinical or clinical development for conditions such as inflammation, cancer and metabolic disease. Mirror DNA and RNA are also used as diagnostic tools, because they retain the features of their natural counterparts but last far longer in blood, tissues and environmental samples. Other possible uses include durable DNA nanomaterials and easily reproduced enzymes as laboratory tools, for medicine, electronics and engineering. Some of these materials have already been produced but have not yet become widely available.

Read carefully, this section is the brief's argument for why a blanket ban would be the wrong instrument. The useful part of the field is the part that is not alive.

What mirror life would require, and why it would not starve

Mirror life means a living, replicating organism, like a bacterium, built entirely from mirror-image components. To function and reproduce it would need mirror versions of DNA to store information, RNA to transmit it, ribosomes to build proteins, and membranes and enzymes to process nutrients.

The question that decides whether it is dangerous is whether it could eat. The brief's answer is yes: mirror organisms could still survive and grow, because both natural and mirror life can use certain basic nutrients, such as oxygen or simple carbon-based substances, that do not have a handedness. They could also be designed to eat nutrients made for the opposite type of life.

So a mirror organism could use basic environmental resources while largely avoiding being attacked by predators, immune systems or rival microbes. Mirror life and natural life would coexist in the same environment but would not interact chemically or biologically, essentially passing through each other without connecting. The brief is careful that this separation does not mean harmlessness: mirror

organisms could survive and spread among humans and in natural environments, potentially causing serious harm to both.

Latest developments

As of today, no reproducing mirror organisms have been created, and many leading experts predict that creating mirror life is at least 10 years away.

However, labs have produced several biological mirror components: mirror DNA, mirror RNA, mirror proteins and mirror enzymes. At least three labs have produced mirror DNA polymerase, the enzyme that enables replication of DNA. Complete mirror ribosomes, metabolisms or membrane systems have not been constructed, all of which would be necessary for self-reproducing mirror life.

This is the brief's progress ledger, and it is also the basis for its red line, because the ribosome is named later as a specific threshold.

Risks and challenges

Leading scientists draw a distinction between the risks of non-reproducing mirror molecules and those of reproducing mirror life. Mirror molecules are already made in many labs and pose few known threats to us or the environment. The creation of mirror life could generate catastrophic risks to ourselves and our planet.

The human health case: because immune recognition and enzymatic breakdown depend on standard molecular handedness, a mirror microbe might evade immune detection, resist common treatments, and avoid predation or infection by natural competitors. Mirror bacteria could therefore be resistant to many human immune responses, which depend on recognising an invasive organism and degrading proteins, and could render known antibiotic treatments largely unusable. In some scenarios, deadly bacteria could replicate and spread globally, with devastating consequences.

The ecological case: mirror bacteria could also infect plants, animals and colonise environments, leading to irreversible ecological disruption. The brief includes the natural brake and then removes its comfort: a mirror organism would grow more slowly than its natural counterpart, because there would be fewer viable nutrients for it to consume, but it could still theoretically spread across the planet. Both animal and plant immune systems rely on recognising invasive cells by their chirality, so mirror bacteria would be resistant to those immune systems and could similarly resist known antimicrobial compounds. The brief's phrase for the result is a highly resilient invasive species with few built-in limitations on its growth.

The detection case is the one that should matter most to anyone building a surveillance system. Detecting and controlling mirror life is especially difficult because almost all diagnostic tests and tools are designed to find regular life forms, not their mirror versions, so these tools probably would not detect mirror DNA or proteins, making it very hard to spot or monitor mirror organisms if they appeared. If a release occurred, early identification with traditional methods would be impossible, increasing the likelihood of wider spread. Once a mirror organism circulated, eliminating it would be extraordinarily difficult, because many existing medical and ecological treatments would be largely ineffective, and its future evolution would be difficult to predict.

These factors also raise security and ethical concerns, since mirror life's inherent biological "invisibility" could be misused by those wishing to do harm, by creating mirror pathogens or invasive mirror species.

And then the dual-use trap, which is the brief's most important structural insight. Scientists are making progress in building the basic parts needed for mirror life, such as special enzymes, cell machinery and genetic material, and this progress is making it much easier and cheaper to assemble a mirror organism than it was before. By simplifying the production of the essential elements of living organisms, even non-replicating mirror biology can accelerate progress

toward mirror life. This is precisely the kind of dual-use risk in the biological sciences identified by the World Health Organization and other experts, and it highlights the need for careful consideration of the risks well before making mirror life becomes feasible for a wide range of actors.

Considerations for the UN system

The brief notes something striking before it gets to policy: very few practical benefits to the creation of mirror life are known, other than the extraordinary scientific accomplishment of generating a living organism solely from synthetic materials.

Distinguishing between the immediate benefits of already-produced mirror molecules and the speculative benefits of mirror life is crucial, especially when considering the risks. In other words, this is an unusually clean risk decision, because almost nothing is being given up.

While mirror life is unlikely to be created within the next 10 years, a growing number of scientists and experts have called for immediate steps to address the risks and identify clear red lines before any breakthrough occurs. UNESCO's International Bioethics Committee and other prominent scientists have called for a global moratorium on creating mirror life.

The four considerations for the UN system are laid out as open questions rather than recommendations.

Whether current international instruments, such as the Cartagena Protocol on Biosafety and the Biological Weapons Convention, and national instruments such as gene-technology regulations, are sufficient to address the risks of technologies that create mirror life, given the unique nature and scale of those risks.

Whether current monitoring and preparedness are sufficient, and if not, what additional capabilities are required. Mirror life may fall outside the detection capabilities and response of current bio-surveillance and public-health systems. This may require the eventual adoption of new international diagnostic standards, laboratory biosafety practices, environmental monitoring or other forms of governance, and experts have suggested that existing capabilities may not be sufficient and that unique solutions may be needed.

What ethical, practical and legal guardrails could safeguard safe research on mirror molecules while ensuring that mirror life is not created. Leading experts have proposed that certain thresholds or red lines are needed for technologies on the pathway to making mirror life, for example preventing the creation of a mirror ribosome. The brief then states the governance gap in one sentence: there is widespread scientific agreement that mirror life

should not be made, but as of today, there is no global process to translate that agreement into governance.

How to encourage ongoing global multi-sectoral collaboration. One step recommended and initiated by a group of prominent scientists has been to launch inclusive multi-disciplinary expert dialogues to develop governance recommendations addressing the societal, ethical, ecological and security implications of mirror biology. Such dialogues could help clarify appropriate technology red lines and acceptable research objectives, articulate shared principles for responsible innovation, establish expectations for transparency in high-capability research, and provide input for policymakers.

What the references tell you

The citations are worth a glance because they show how new this consensus is. The technical basis is Adamala and colleagues' 2024 *Science* paper "Confronting risks of mirror life" and their 2024 "Technical Report on Mirror Bacteria: Feasibility and Risks". The governance activity is from 2025: a Paris Conference on Risks from Mirror Life, a UK Government Office for Science analysis, UNESCO's International Bioethics Committee draft report, and a "Spirit of Asilomar" entreaty signed by 96 conference participants stating that mirror life should not be created unless future research convincingly demonstrates it would not pose severe risks. The brief

also records that two major science philanthropies, Renaissance Philanthropy and the Alfred P. Sloan Foundation, have publicly committed to not funding research that aims to make mirror life.

Words to know

- **Chirality (handedness)** — the property of a molecule existing in two mirror-image forms that cannot be superimposed, like left and right hands. Everything in the brief follows from it.
- **Mirror biology versus mirror life** — mirror biology makes individual molecules with flipped handedness and already exists; mirror life would be a self-replicating organism made entirely of them and does not. The brief's central argument is that these two must not be governed as one thing.
- **Spiegelmer** — a short mirror-image RNA strand that binds a disease-related protein while avoiding immune detection and enzymatic degradation. Several are in preclinical or clinical development, and they are the brief's best example of a real benefit.
- **Ribosome** — the cellular machine that builds proteins. It matters here because a mirror ribosome has not been built, all mirror life would need one, and experts propose preventing its creation as a red line.
- **DNA polymerase** — the enzyme that enables DNA replication. At least three labs have made a mirror

version, which is why the brief treats the timeline as moving.

- **Generalist heterotroph** — an organism that can consume a wide range of organic material rather than one specific food source. This is the property that would stop mirror bacteria being self-limiting, since they would find nutrients in hosts and the environment.
- **Biocontainment** — designing an organism so it cannot survive outside controlled conditions. The brief's concern is that mirror bacteria would not be intrinsically biocontained.
- **Dual-use risk** — where the same work serves benefit and harm. Here it has an unusual shape: the harmless, useful molecular work is itself the route to the dangerous organism.
- **Cartagena Protocol on Biosafety** — an international agreement on the transboundary movement of living modified organisms. Named as one of the existing instruments that may or may not be adequate.

Numbers in this reading

- Six summary points; the brief runs to 6 pages.
- Mirror life has not been created and is likely a decade away; many leading experts predict at least 10 years.
- At least three labs have produced mirror DNA polymerase.

- Complete mirror ribosomes, metabolisms and membrane systems have not been constructed, and all are necessary for self-reproducing mirror life.
- Very few practical benefits to creating mirror life are known.
- 96 conference participants signed the "Risks from Mirror Life" Spirit of Asilomar entreaty (4 April 2025).
- Two major science philanthropies, Renaissance Philanthropy and the Alfred P. Sloan Foundation, have publicly committed to not funding research aiming to make mirror life.

What the author is unsure about

The brief's language is carefully conditional throughout, which is itself the point: this is a risk assessment of something that does not exist. Mirror life is "still hypothetical" and "likely" a decade away. A mirror microbe "might" evade immune detection; deadly mirror bacteria could spread globally "in some scenarios"; a mirror organism "could still theoretically" spread across the planet; diagnostic tools "probably would not" detect mirror DNA or proteins. Its future evolution would be difficult to predict. The open questions about whether existing treaties, monitoring and preparedness are sufficient are posed as questions, with the hedge that experts "have suggested" existing capabilities may not be enough. And the standing disclaimer says the views are the

Board's independent assessment rather than those of the UN or of the scientific community writ large. What the brief does not hedge: mirror life has not been created; no mirror ribosome exists; there is widespread scientific agreement that mirror life should not be made; and there is no global process to turn that agreement into governance.

For the discussion

- A point worth making: this is the rare risk decision with almost nothing on the other side of the ledger. The brief says very few practical benefits to creating mirror life are known beyond the achievement itself, while the useful mirror therapeutics and diagnostics do not require a living organism at all.
- A tension worth raising: the dual-use shape here is inverted from the usual case. The benign molecular research is what makes the dangerous organism cheaper and easier, so a governance regime that permits the benefits inevitably shortens the path to the harm.
- A question worth asking: she is certified in Frontier AI Governance and AGI Strategy at BlueDot Impact (October 2026), so she has just spent a course on exactly this problem, which is drawing a red line before a capability exists. The question to put to the group is whether "do not build a mirror ribosome" is a workable red line or only a slogan,

given that nobody is currently positioned to verify it.

1.4-1 • CORE • COURSE TIME 30 MIN

Pandemic prevention as fire-fighting

Richard Williamson • 2022

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

Fire has almost disappeared as a cause of death in the developed world, and it happened through materials, codes, alarms and insurance rather than braver firefighters, so Williamson argues the same four-part strategy (prevention, detection, passive protection, active protection) is how we tame pandemics.

Average summary

Williamson opens with smallpox and then sets it aside. In November 1977 a smallpox eradication team arrived in the Somalian port city of Merca and vaccinated 54,777 people in two weeks, going house to house and setting up checkpoints to vaccinate everyone entering and leaving. They were responding to Ali Maow Maalin, a 23-year-old hospital cook who had taken a car ride of no more than 15 minutes with two children infected with smallpox, after being asked for directions. His fleeting exposure was missed by

contact tracers; ten days later he got sick. In May 1980 the 33rd World Health Assembly declared smallpox eradicated, and the last push cost only about one billion dollars in today's currency.

But smallpox was unusually eradicable: a human disease with no animal reservoir, distinctive symptoms, no asymptomatic transmission, and a DNA genome with a much slower mutation rate than many viruses. Most natural diseases of pandemic potential differ in one or more of these ways, and that is before considering engineered ones. So he looks elsewhere for a model, and finds **fire**.

The comparison does real work. Boston 1760, New York 1776 and 1835, Chicago 1871, Baltimore 1904: all events called the Great Fire. The total cost of fire in the United States today is on the order of two percent of GDP, but actual losses are only 17 percent of that; the remaining 83 percent, \$273 billion in 2014, is prevention and protection, and only \$40 billion of that is fire departments. That is like funding an entire Apollo lunar programme, about \$280 billion in today's dollars, every single year. Meanwhile the deadliest fire in US history killed around 3,000 people, all civilian fire deaths from 1977 to 2020 came to approximately 200,000, and COVID-19 in the United States alone is around 350 times worse than the worst ever fire and five times worse than all fires in the last 45 years combined.

He then runs his four categories. Active protection: just as the four classes of fire need different extinguishers, the roughly 26 families of viruses known to infect humans differ, and for around half of them we have no licensed vaccine for any virus in the family. Detection: fire alarms cut reporting time by about 20 minutes and halve the risk of fire death, while SARS-CoV-2 circulated two to three months before its genome was posted; PCR needs a known target, so it has to be complemented by target-agnostic metagenomic sequencing, which is currently too expensive and whose data nobody yet knows how to analyse. Passive protection: this is where he is most optimistic, covering HEPA filtration, ventilation, far-UVC light below 235 nanometres, and next-generation PPE. Prevention: the leading cause of fire is human behaviour, and the biological equivalents are gain-of-function research, lab leaks and deliberate misuse.

Why it's assigned

This is the Core reading for 1.4 and the direct input to the day's exercise, which asks her to imagine a pandemic-proof 2050. It is also where the course's four-layer structure visibly comes from: prevention becomes Unit 2, detection becomes Unit 3, passive protection becomes Unit 4, and active protection becomes Unit 5. The fire analogy is not decoration; it is the course's argument that a hazard can go from

civilisation-shaping to almost forgotten, and that this happened through boring infrastructure rather than heroism.

In depth

Why smallpox is the wrong model

The Merca story is told for what it concedes. The team vaccinated 54,777 people in two weeks; by 1976 Somalia was the last country with cases still being detected. Had Maalin been identified and vaccinated at once, the outbreak could have ended with the two children in the car. Searches continued into 1979 and found nothing, and in May 1980 the 33rd World Health Assembly declared smallpox eradicated. The virus survives only in lab samples, including the one that claimed smallpox's last victim, and in the knowledge of its genetic code; the reading does not name that victim or the accident.

Then the deflation. Smallpox was unusually **eradicable**: a human disease with no animal reservoir, distinctive symptoms, no asymptomatic transmission, and a DNA genome that mutates far more slowly than many viruses. Most pandemic-potential diseases differ in one or more of these ways, before you consider engineered ones. So we must look elsewhere for inspiration.

The gift of Prometheus

He lists the North American events named Great Fire: Boston 1760, New York City 1776, New Orleans 1788 and 1794, New York 1835 and 1845, Pittsburgh 1845, Toronto 1849, Montreal 1852, Portland 1866, Chicago 1871, Boston 1872, Vancouver 1886, Baltimore 1904, Toronto 1904. The Great Fire of London in 1666 produced private fire insurance and the first organised fire protection, run by those insurers.

The cost accounting is the centre of the essay. Fire costs the United States on the order of two percent of GDP, but not because there is still much fire. Actual losses are only 17 percent of that, mostly the statistical cost of death and injury at a value of a statistical life of \$9.6 million. The remaining 83 percent is protection and prevention: \$273 billion in 2014, of which only \$40 billion is fire departments, with far more spent making buildings and products firesafe. He calls these crude high-level estimates, noting that making nonresidential buildings firesafe is assumed to cost construction expenditure times 12 percent. The scale: like funding an entire Apollo lunar programme, about \$280 billion in today's dollars, every single year. And it is worth it, because the cities used to burn.

What if fire, but worse

The numbers close the gap. The deadliest US fire, triggered by the 1906 San Francisco earthquake, killed around 3,000 people, and all US civilian fire

deaths between 1977 and 2020 come to approximately 200,000. On US deaths alone, COVID-19 is around 350 times worse than the worst ever fire and five times worse than all fires in the last 45 years combined, at a cost of around \$16 trillion in the US, about 50 years' worth of fire's annual cost. Since 1900, no fewer than five pandemics have killed 100,000 or more in the United States alone: three influenza pandemics, COVID-19 and HIV/AIDS. His framework: prevention, detection, passive protection and active protection, because pandemics, like fire, are not primarily about more responders.

Active protection

The four broad types of fire (ordinary combustibles, flammable liquids and gases, metal, electrical) need different technologies, which is why chemical foam and carbon dioxide extinguishers mattered: they allowed effective suppression of flammable-liquid fires for the first time.

Biology is less kind to us than chemistry.

Approximately 26 families of viruses are known to infect humans, and the differences between families matter as much as the classes of fire, though viruses within a family share structural features so some drugs work across one. Two examples: in the monkeypox outbreak current at the time of writing, the Jynneos vaccine and the antiviral tecovirimat were originally developed against a potential bioterrorist

smallpox attack; and the annual flu shot is a new formulation each year on the same underlying technology. The COVID-19 case shows why prior work matters: a crucial piece of the vaccine puzzle was meeting the coronavirus spike protein in the shape it has before it unlocks the door into our cells, and that prefusion spike protein was worked out while designing a MERS vaccine.

His asks: for around half of the virus families that infect humans we have no licensed vaccine for any virus in the family, and that must be the starting point. Use a prototype pathogen approach for antivirals and monoclonal antibodies for at least one virus per family; aim for broad-spectrum vaccines and antibodies for whole families and antivirals across families; and accelerate authorisation for monoclonal antibodies, learning from the expedited pathway for mRNA vaccines.

Detection

Fire alarms cut reporting time by an average of around 20 minutes in the United States, against a UK emergency response averaging seven to eight minutes and a US target of 90 percent of incidents with the first engine in just over five minutes. His point is where the leverage sits: huge effort goes into shaving seconds off response, and all of it pales beside starting earlier. Hence a 2016 meta-analysis finding smoke alarms halve the risk of fire death, and

a Canadian study finding a 71 percent cut in the risk of a fire spreading beyond the room of origin.

For pandemics the timeline is the indictment. SARS-CoV-2's first complete genomic sequence was posted on virological.org on 10 January 2020, Wuhan reported the first death the next day, and the virus is thought to have been circulating two to three months earlier. A virus doubling weekly infects a thousand times more people per day after ten weeks, and doubling every 2.5 days, as estimated in Hubei in January and February 2020, turns one new infection per day into millions in under two months.

The technical problem is specific. PCR amplifies sequences but needs a known target called a template, so targeted testing cannot find a new threat however cheap it gets. It must be complemented by target-agnostic metagenomic sequencing, which reads everything in a sample and could flag threats from environmental samples before anyone is sick enough to seek urgent care. He does not oversell it: possible now in theory, but the sequencing would be prohibitively expensive and nobody yet knows how to analyse the data, which the Nucleic Acid Observatory at the MIT Media Lab is working on. His immediate complaint is that we fail to use what we have: poliovirus is found in wastewater worldwide using PCR, and in New York they only started looking after the first case of paralysis in an unvaccinated patient.

Passive protection

This is where he is most optimistic, because we have engineered the whole urban environment against the spread of fire through materials, codes and enforcement, at real political cost: after the Great Fire of Chicago in 1871 restrictions on wood were watered down by working-class opposition to higher home costs, and failure to enforce can end in tragedy like the 72 lives lost at Grenfell Tower.

The near-term biological measures: HEPA filters can already remove SARS-CoV-2 from indoor air on a hospital ward and ventilation is extremely effective against airborne pathogens, so funding trials in high-risk buildings is a no-brainer. Even outside pandemics, hospital-acquired infections are a significant burden with the airborne share estimated in the region of ten percent, so reducing them alone could be worth billions.

The emerging technology is far-UVC. UVC is a potent antimicrobial, but wavelengths over 235 nanometres penetrate skin to the DNA in our cells and can cause cancer; below 235 nanometres it is absorbed by even a small layer of protein, so photons are blocked by the outer layer of dead skin cells while staying effective against airborne microbes. Then the open questions, the most responsible paragraph in the essay: effects on common materials, cosmetics and surfaces; interaction with existing ventilation; effectiveness

against larger respiratory droplets; whether sources emit above 235 nanometres as a by-product; trade-offs between sub-235-nanometre wavelengths, since most studies use 222 nanometres; and edge cases such as an open wound. It is still the most promising new candidate for a pandemic firewall.

We can also wear a firewall. Current PPE is uncomfortable, unattractive, impractical, insufficiently reusable and imperfectly protective; next-generation PPE needs innovation in materials, design and engineering, adding features such as neutralising pathogens with antimicrobial materials like copper and detecting exposures. Stockpiling has outsize impact if an outbreak is suppressed early, and otherwise PPE must be rapidly producible from widely available materials, which he thinks needs public and philanthropic support through advance purchase commitments as pioneered in vaccines.

Prevention

The number one cause of fire is human behaviour; when deliberate we call it arson, but mostly it is the innocent, occasionally negligent, pursuit of work and play. The leading cause of US home fires is cooking, the main factor is unattended equipment, and more than a quarter of those who died from cooking fires were asleep. His framing: with comprehensive protection and detection, unsafe fire behaviour does not imperil whole cities, so we need not restrict deep

fat fryers; were conflagration still realistic, our risk tolerance would be far lower. So tolerance for risky research should depend on how good our defences are.

He works the gain-of-function case concretely. H5N1 Influenza A has a fatality rate in excess of 50 percent among identified cases and is rare in humans because it does not transmit efficiently between them; in the year of writing, tens of millions of birds were culled. Researchers tested how likely human-to-human transmission is using an animal model, here ferrets, and created a weakened H5N1 that was ferret-to-ferret transmissible, which the wild virus could not do. He is careful with terms: gain of function is imprecise, since new properties can be innocuous, like green fluorescent protein to make things glow; what matters is dangerous functions, what the NIH calls enhanced potential pandemic pathogens, inside the umbrella of dual-use research of concern. Advocates say we learned that evolution of H5N1 toward transmission in a useful human model is possible; detractors say the benefits are speculative and were obtained by creating a virus that could potentially transmit between humans. His analogy: a fire drill with an actual fire may be a better drill, but does the risk justify the reward?

On lab leaks he is scrupulous. The evidence seems to point away from a Wuhan Institute of Virology origin

for COVID-19, but a pandemic from an accidental breach of a high-containment laboratory is scientifically credible, with precedent even from BSL-4 labs: the 2007 UK foot-and-mouth outbreak was traced with high likelihood to a leaky waste pipe at a joint BSL-4 lab and vaccine site, and after relicensing the virus leaked again into the drainage system. More striking, he argues we have probably already had the first human pandemic caused by escape from a lab, because the 1977 Russian flu is so genetically similar to ancestral H1N1 strains, given influenza's normal mutation rate, that a lab sample is the most plausible origin.

Proliferation, or biological arson

Williamson is much more confident we will keep seeing natural pandemics than engineered ones, but engineered ones could be far deadlier, and mass civilian death from a bioweapons facility accident has already happened at least once: the 1979 release of weaponized anthrax spores from a Soviet plant is estimated to have killed at least 66 people, limited only because anthrax is not contagious.

On states he focuses on intent rather than capability. Many nuclear arms control successes were about making pursuit unattractive rather than impossible, and where perceived benefits outweigh costs states proceed, as North Korea has despite being desperately poor with low state capacity. The

analogue to the Nuclear Non-Proliferation Treaty is the Biological Weapons Convention, with 183 signatories including all major powers, but it lacks enforcement and has far less institutional, political and financial support than the nuclear and chemical regimes. His line for scale: the BWC runs on a budget smaller than that of a single McDonald's restaurant. The Soviet Union signed it and secretly continued a vast programme, and the extent of such programmes today is impossible to know for sure.

The short-term worry is non-state capability. Thanks to the increasing ubiquity of DNA synthesis technology, the people capable of assembling an influenza virus from publicly available genetic sequences number in the tens of thousands, and in an act that might plausibly have caused more harm than good, researchers proved they could reconstitute horsepox, an extinct virus similar to smallpox, for a mere \$100,000. If individuals could order highly enriched uranium online we would all call that a threat to global security, so we must constrain the biological equivalent. The chokepoint is synthesis screening, and providers of synthetic DNA are not legally required to screen for dangerous sequences, though many do voluntarily. The Nuclear Threat Initiative is launching the International Biosecurity and Biosafety Initiative for Science (IBBIS), whose mission includes preventing misuse of DNA synthesis technology, and the Secure DNA Project aims to give companies

worldwide free, fast and secure screening of orders. These efforts show promise, and it remains to be seen whether they will universalize screening.

The will to act, and the ending

We did not solve the Great Fire through more and braver firefighters: the transformation in the 160 years since James Braidwood, author of the first English textbook on firefighting, died at the Tooley Street warehouse fire of 1861, came from innovators, engineers, campaigners, regulators and legislators, and from the will to act.

The policy history is deflating. The strategies here are inspired by the Apollo Program for Biodefense, a bipartisan initiative the US government came tantalizingly close to embarking upon, and the American Pandemic Preparedness plan proposed the previous year would have authorized \$88.2 billion over five years, less than one percent of COVID-19's costs. He says that is not nearly enough but would have been a start. Then one sentence alone: Congress did not take up the call. He notes a new window in the recently released National Biodefense Strategy and Implementation Plan, the clearest plan yet, assigning actions to the relevant agencies. Then: how many more biological Great Fires do we need in order to act?

He closes back in Merca. Ali Maow Maalin recovered and later worked to eradicate polio from Somalia. One of the children in the car, a six-year-old girl named Habiba Nur Ali, became the last person to lose her life to a wild smallpox infection. In 2013 polio reemerged in Somalia, and Ali was vaccinating in his hometown when he developed a sudden fever; days later he died in hospital of malaria. Then the thesis of the ending: the end of pandemics will not look like the end of smallpox. There will never be a last outbreak of a disease of pandemic potential, just as there will never be a last fire, never a last victim, never a single moment where we can say we did it. Ali could have thought himself fortunate and lived an ordinary life, but he stepped up and carried the fight to the next frontier. He had the will to act. The question, Williamson ends, is whether we do.

Words to know

- **Animal reservoir** — a non-human population in which a pathogen persists. Smallpox had none, which is why it could be eradicated, and most pandemic-potential diseases do, which is why they cannot.
- **Prototype pathogen approach** — picking one representative virus from each viral family and developing vaccines, antivirals and antibodies for it, so the groundwork is done before the family produces a pandemic. It is the active-protection

equivalent of stocking the right class of extinguisher.

- **Metagenomic sequencing** — reading every piece of genetic material in a sample rather than testing for one known target. The only way to find something you are not looking for, and currently too expensive to do at scale.
- **Template (in PCR)** — the known gene sequence a PCR test amplifies. Without one there is no test, which is the single clearest reason targeted testing cannot be an early-warning system.
- **Far-UVC** — ultraviolet light below 235 nanometres, absorbed by a thin protein layer so it does not reach the DNA in our cells, while still destroying airborne microbes. Most studies have been done at 222 nanometres.
- **Nosocomial infection** — an infection acquired in a healthcare setting. Williamson uses it to argue ventilation and filtration pay for themselves outside pandemics.
- **Enhanced potential pandemic pathogen (ePPP)** — the NIH's term for a pathogen deliberately given dangerous new properties. Williamson prefers it to "gain of function", which he calls imprecise.
- **Dual-use research of concern** — research with potential to do harm as well as good; the umbrella category that ePPP work sits inside.
- **Prefusion spike protein** — the shape the coronavirus spike has before it enters a cell, which

the immune system needs to learn. It was worked out for a MERS vaccine, and reusing it is why COVID-19 vaccines came as fast as they did.

- **DNA synthesis screening** — checking ordered genetic sequences against dangerous ones before manufacture. Not legally required of providers, though many do it voluntarily.

Numbers in this reading

- Merca, November 1977: 54,777 people vaccinated in two weeks; Ali Maow Maalin was 23; the car ride was no longer than 15 minutes; he fell sick ten days later; searches continued into 1979; eradication declared May 1980 by the 33rd World Health Assembly; the last push cost about one billion dollars in today's currency.
- Total cost of fire in the US: on the order of two percent of GDP. Actual losses: 17 percent of that. Protection and prevention: 83 percent, or \$273 billion in 2014, of which only \$40 billion is fire departments. Value of a statistical life used: \$9.6 million. Nonresidential firesafe cost assumed at construction expenditure times 12 percent. Apollo comparison: about \$280 billion in today's dollars.
- Deadliest US fire (1906 San Francisco earthquake): around 3,000 deaths. All US civilian fire deaths 1977 to 2020: approximately 200,000. COVID-19 in the US: around 350 times worse than the worst fire and five times worse than all fires in the last 45

years combined; total cost around \$16 trillion in the US alone, about 50 years of the annual cost of fire. Five pandemics have killed 100,000 or more in the US since 1900.

- Approximately 26 families of viruses are known to infect humans; for around half we have no licensed vaccine for any virus in the family.
- Fire alarms cut reporting time by about 20 minutes; UK emergency response averages seven to eight minutes; the US target is 90 percent of incidents with the first engine in just over five minutes. A 2016 meta-analysis found smoke alarms halve the risk of fire death; a Canadian study found a 71 percent reduction in the risk of a fire spreading beyond the room of origin.
- SARS-CoV-2's first complete genomic sequence posted 10 January 2020; first Wuhan death reported the next day; circulating about two to three months earlier. Doubling every week means a thousandfold increase after ten weeks; doubling every 2.5 days, as estimated in Hubei in January and February 2020, turns one new infection per day into millions in less than two months.
- About 1,000 home structure fires a day in the USA. Grenfell Tower: 72 lives lost. Triangle Shirtwaist Factory, 1911: 146 dead. Winecoff Hotel, 1946: 119 dead. Great Fire of Chicago 1871; Great Fire of Boston 1872; Lübeck's wood restriction in the thirteenth century.

- Airborne share of nosocomial infections: in the region of ten percent.
- Far-UVC: the safety boundary is 235 nanometres; most studies have been done at 222 nanometres.
- H5N1: fatality rate in excess of 50 percent among identified cases; tens of millions of birds culled in the year of writing.
- Sverdlovsk, 1979: at least 66 people estimated killed. The 1977 Russian flu pandemic's most plausible origin is a lab sample.
- Biological Weapons Convention: 183 signatories; run on a budget smaller than a single McDonald's restaurant.
- People capable of assembling an influenza virus from publicly available sequences: tens of thousands. Horsepox reconstituted for a mere \$100,000.
- American Pandemic Preparedness plan: \$88.2 billion over five years, less than one percent of COVID-19's costs. Congress did not take it up.
- Habiba Nur Ali was six years old; polio reemerged in Somalia in 2013; Ali Maow Maalin died of malaria.

What the author is unsure about

Williamson flags his own weak points more often than most advocates. He calls the fire cost figures crude high-level estimates and shows one of the

assumptions behind them. He says a global metagenomic early-warning system is theoretically possible at the current state of technology but prohibitively expensive, and that we do not yet know how we would analyse the data it would produce. His far-UVC section is a list of open questions: effects on materials, cosmetics and surfaces; interaction with existing ventilation; effectiveness against larger respiratory droplets; whether deployed sources emit above 235 nanometres as a by-product; trade-offs between sub-235-nanometre wavelengths; and edge cases like open wounds. He says the evidence seems to point away from a Wuhan Institute of Virology origin for COVID-19 while maintaining that lab-leak pandemics are scientifically credible, and he marks the 1977 Russian flu lab origin as the most plausible explanation rather than a proven one. He says he is much more confident we will continue to see natural pandemics than engineered ones. He says the extent of continuing bioweapons programmes in Russia or anywhere else is impossible to know for sure. He says it remains to be seen whether the voluntary synthesis-screening efforts will universalize. And he presents the gain-of-function debate as two positions with a question attached rather than a conclusion.

For the discussion

- A point worth making: the fire comparison's real punchline is the budget split. Of what the United

States spends on fire, only about 15 percent of the protection spending goes to fire departments; the rest is materials, codes and inspection. If pandemic defence follows the same shape, most of the money is not medical.

- A tension worth raising: he argues our tolerance for risky research should depend on how good our defences are, which is a strong claim in both directions. It justifies gain-of-function restrictions now, and it also implies the restrictions should loosen once passive protection is in place.

1.6-1 • CORE • COURSE TIME 180 MIN

Biosecurity Really: A Strategy for Victory

Drew Endy, Sarah Moront, Vassilis Andrea Alexopoulos, Raj Patel, Rhea Jain, Britney Bennett • 2025

[Read the original paper →](#) • the course assigns: from the preface, the chapter 1 conclusion, the blueprint introduction, the final chapter "Biosecurity Victory", and every section heading

A Hoover Institution report that refuses to treat biosecurity as damage limitation: it argues biology itself is the best defence against biology, that the United States could spend no more than \$40 billion a year to get there, and that the sequence of actions matters more than the list.

Average summary

From the preface, the chapter 1 conclusion, the policy blueprint's introduction, the final chapter and every section heading. Published October 2025 by Hoover's Bio-Strategies and Leadership programme; the preface is signed by Drew Endy, Stanford, August 2025.

The preface works by contrast. Two versions of the same Sunday family dinner. In the first, a teenager has a runny nose and nobody knows what it is, Dad has been sick three times already this school year, and Mum calls in sick on Tuesday with a fever. In the second, the son knows his sniffles are **HRV-B14**, a relatively mild human rhinovirus strain; Mum took a multivalent common cold vaccine earlier in the year; Dad's school is warned in advance when viruses are circulating and modern air-treatment systems stop them spreading in classrooms; and "smart" facial tissues give the family confidence the boy is no longer infectious. Which dinner happens depends on what we make real over the next three, five and fifteen years.

The report has five chapters. Chapter 1, Worst-Case Scenarios, lays out eight trends making biosecurity worse: proliferation of capacities that make creating or deploying biological threats easier; routinized creation of novel toxins or pathogens; nation-states reembracing biological weapons programmes; the

United States no longer leading the world in biotechnology; erosion of public health systems; urbanization, globalization, travel and agriculture amplifying risk; degraded biodiversity and ecosystems; and societal trends. Chapter 2 covers today's pillars of biosecurity and emerging biotechnologies for surveillance, vaccines, diagnostics and treatments. Chapter 3, *How Far Are We Willing to Go?*, evaluates beyond-frontier biotechnologies under four headings: environmental interventions, gain of function, nonheritable interventions and heritable interventions.

Chapter 4 is the blueprint: nine policy recommendations in three stages, with Stage 1 inside the next thousand days, Stage 2 through 2030 and Stage 3 organising capacities for by or before 2040. The nine are: seeing behind the molecular curtain; distrust and bioaudit; building a governance regime for dangerous pathogens; working with allies and adversaries; using biology to secure biology; governing at the speed of biology; strengthening our digital biodefenses; securing our air; and promoting bioliteracy and rewarding biosecurity leadership.

Chapter 5 states the premise plainly: we should expect an era of near unlimited-access biotechnology, in which most people in most places will eventually have the capacity to make nearly any toxin or pathogen, almost all will have zero interest in doing

so, and some eventually will. Its warning about states is that eroding norms may nudge nations toward what the report calls Stupidly Assured Destruction. Its costing: no more than \$40 billion annually to the US federal government, generating up to \$1 trillion annually in direct and indirect public and economic benefits, against the \$95 billion per year the United States is expected to spend on its nuclear arsenal over the next ten years.

Why it's assigned

It is one of three long optional overviews on the 1.6 page, and it is the most forward-leaning thing the course offers. Where the rest of Unit 1 establishes the problem, this report insists on a strategy with a win condition, a budget and an order of operations, and its claim that stabilizing actions such as biological intelligence must happen first is the sharpest answer in the whole unit to the day's third question about what a pandemic-proof world would look like and how we get there. Its opening scene also shows the thing the course keeps pointing at, which is that a pandemic-proof world is mostly made of ordinary infrastructure: known diagnoses, treated air, advance warning.

Words to know

- **BIOINT (biological intelligence)** — the report's name for unified, cross-agency, multimodality

biosurveillance whose mission is that we are never surprised by biological threats. It is recommendation 1 and the report says it must come first.

- **BIONET** — the report's term for distributed biomanufacturing resilience, built on the ARPANET analogy: a network so that production can continue when any one node fails.
- **Biotrust** — the report's term for the trust relationships between scientists, policymakers and the public that governance depends on. It pairs with "bioaudit", the verification that makes trust reasonable.
- **Stupidly Assured Destruction (S.A.D.)** — the report's label for nations dangerously expanding their arsenals with modern biological capabilities, in deliberate contrast to the Cold War's mutually assured destruction.
- **Beyond-frontier biotechnology** — a technology not yet feasible but close enough to need an ethical decision now. Chapter 3's job is to evaluate these before they arrive.
- **Technology readiness level (TRL)** — a standard scale for how close a technology is to deployment; the report pairs it with its own society readiness levels, running from warmly welcomed and widely believed down through significant objection.
- **Bioliteracy** — a general public understanding of and ability to engage with biology. The report treats

it as a defence, not an educational nicety.

- **Gain of function** — research giving an organism new or enhanced properties. Endy's preface says policy here failed to adapt to changing technology and created ambiguity about pandemic origins.
- **Multivalent vaccine** — one that protects against several strains or pathogens at once. The common cold vaccine in the second dinner is the report's image of what this unlocks.

Numbers in this reading

- Published October 2025; preface dated August 2025.
- The next thousand days is the report's Stage 1 window; Stage 2 extends through 2030; Stage 3 organises capacities for by or before 2040.
- Nine policy recommendations across five chapters.
- Cost: no more than \$40 billion annually to the US federal government, against up to \$1 trillion annually in direct and indirect public and economic benefits.
- Comparison: \$95 billion per year is what the United States is expected to spend to operate, maintain and upgrade its nuclear arsenal over the next ten years.
- The named rhinovirus strain in the opening scene is HRV-B14.

- Recommendation 1 has four parts, 1A through 1D, flagged Stage 1, Stage 1, Stage 2 and Stage 3.

What the author is unsure about

The report is explicit that its recommendations are judgements. The preface says the opinions, examples, conclusions and policy recommendations, along with any errors or misjudgements, are the authors' own, and that one of the three challenges is to understand changing risks without becoming consumed by unfounded hypotheticals, naming "too many might and could" as the failure mode. The blueprint says the recommendations are written with the United States in mind but can be adopted elsewhere as appropriate, and that some specific details are meant to be illustrative rather than prescriptive, since options exist for accomplishing certain goals. Endy records that at the start of the project he was unsure whether a calm and stable biosecurity victory might be possible, and that he is now optimistic, which marks the optimism as a conclusion he reached rather than a premise. Stage 3 is defined as capacities that may never be needed. And the cost and benefit figures are given as expectations ("we expect that... would cost", "generating up to"), not as calculations the reader can check from the text.

For the discussion

- A point worth making: this report's distinctive claim is about sequence, not content. Everyone's list of biosecurity measures looks similar; the argument that stabilizing actions such as biological intelligence must come first, because the others depend on knowing what is happening, is the part worth arguing about.
- A tension worth raising: the authors accept that most people in most places will eventually be able to make nearly any toxin or pathogen, and then propose governance, audit and licensing. If near-unlimited access is the premise, defences that work regardless of intent have to carry almost all the weight.
- A question worth asking: the preface comes from someone who sat on the boards whose policies he says failed. What does the report think went wrong in the gain-of-function governance he helped run, and what makes the new recommendations less likely to age the same way?

1.6-2 · CORE · COURSE TIME 120 MIN

Global Catastrophic Biological Risks A Guide for Philanthropists

Christian Ruhl · 2023

[Read the original paper →](#) · the course assigns: from the summary, the note on hazardous information, the conclusion,

the spending and neglectedness sections, and every section heading

Founders Pledge's guide argues that the worst biological catastrophes are both disproportionately worse than ordinary pandemics and disproportionately neglected, and gives philanthropists a set of heuristics for finding the few funding opportunities that address that gap.

Average summary

From the summary, the note on hazardous information, the conclusion, the two sections on societal spending and philanthropic neglectedness, and every section heading. Written by Christian Ruhl for Founders Pledge, October 2023, using semi-structured interviews and analysis of funding and pandemic preparedness data.

Three claims carry the report. The expected cost of a biological catastrophe is immense: high-consequence events like engineered pandemics are among the biggest threats to human life and modern civilization this century, and the case for preparedness stays strong even without placing special value on existential risk. The threat landscape is rapidly changing, because advances in the life sciences may increase risk both by providing new capabilities and by putting them in reckless and malicious hands,

which also increases uncertainty about the origin and characteristics of future threats. And societal spending remains misallocated even after COVID-19: governments and traditional philanthropists spend billions on public health and health security but **disproportionately neglect the most high-consequence threats.**

The heuristics are the usable part. Focus on worst-case scenarios, since outbreaks follow very heavy-tailed distributions where the worst pandemics outweigh many smaller ones combined, which in practice often means prioritizing engineered threats, because nature is not the worst bioterrorist. Pursue pathogen- and threat-agnostic approaches rather than guessing where the next pandemic comes from. Fund interventions robust across the whole spectrum of risk, which often means prioritizing prevention over response. Leverage existing societal resources through advocacy, mostly aimed at government. Prioritize interventions with near-term positive externalities that can win broad political support. And prioritize low dual-use potential and high offense-defense distinguishability, avoiding information hazards.

His funding opportunities: GCBR field-building and policy advocacy; safeguards against proliferation and misuse of key technologies and information, especially AI misuse and DNA synthesis screening;

transmission-blocking interventions, including sterilizing technology, germicidal lights for indoor biological air quality and mechanisms to quickly decontaminate PPE; pandemic-proof personal protective equipment, a term he credits to Kevin Esvelt; pathogen-agnostic early warning, including metagenomic sequencing; and platform technologies for medical countermeasures that can rapidly pivot for novel threats.

On the spending evidence he is careful rather than rhetorical. US global health security funding is about \$1.5 billion in fiscal year 2023, much of it for endemic diseases. The Biden administration touted a historic \$88.2 billion request in the 2023 budget, which was requested rather than appropriated. Between 2001 and 2014, only about 21% of biodefense funding was strictly biodefense rather than multi-use. Excluding COVID response, the FY 2023 budget requested about \$11 billion in discretionary biodefense funding, and only a small slice is plausibly relevant to worst-case scenarios. On the philanthropic side he names Open Philanthropy, Effective Giving and entities associated with Effective Ventures, and reports that excluding direct response and HIV-related funding, total philanthropic funding remains in the tens of millions.

Why it's assigned

It is the second of the three long optional overviews, and it is the one that answers a question the rest of Unit 1 leaves open: if all of this is true, who pays, and for what? Reading it alongside the core readings is useful because it independently arrives at the same shortlist the course is built around (prevention and DNA synthesis screening in Unit 2, pathogen-agnostic early warning in Unit 3, transmission-blocking and P4E in Unit 4, platform countermeasures in Unit 5). It is also the clearest statement in the unit of the discipline that handling dangerous information requires, which Unit 6 takes up.

Words to know

- **GCBR (global catastrophic biological risk)** — a biological event severe enough to threaten human life and modern civilization, as distinct from an ordinary pandemic. The report's whole argument depends on keeping the two apart.
- **Heavy-tailed distribution** — one where the rare extreme events dominate the total, so the worst pandemic outweighs many smaller ones combined. This is the statistical reason for focusing on worst cases.
- **Impact multiplier** — Ruhl's term for a quasi-quantitative heuristic that helps identify where an extra pound or dollar does the most good. The report's six heuristics are all of this kind.

- **Threat-agnostic (pathogen-agnostic)** — an approach that does not depend on knowing which pathogen is coming. His argument is that it is both higher-impact and safer, because it avoids the information hazards of naming specific threats.
- **Offense-defense distinguishability** — whether an investment can be told apart from an offensive capability by an outside observer. Low distinguishability feeds security dilemmas, where defensive moves provoke rivals.
- **Security dilemma** — the dynamic where one state's defensive preparation looks threatening and triggers the behaviour it was meant to deter. The report's grantmaker checklist exists partly to screen for this.
- **Information hazard** — information that increases risk simply by being spread. The report sets out its own practices for avoiding it and asks funders to do the same.
- **P4E (pandemic-proof personal protective equipment)** — protective equipment good enough to be trusted in a worst-case pandemic, a term the report credits to Kevin Esvelt.
- **Platform technology** — a manufacturing and development approach that can be re-aimed at a new pathogen rather than rebuilt, as mRNA vaccine production can.
- **Deterrence by denial** — preventing an attack by making it impossible to succeed rather than by

threatening retaliation. It appears in the conclusion's image of bio-bunkers.

Numbers in this reading

- US global health security funding: about \$1.5 billion in fiscal year 2023.
- The Biden administration's 2023 budget included a historic \$88.2 billion request for mandatory funding to prepare for future biological threats, which the report notes was requested rather than appropriated.
- Between 2001 and 2014, only about 21% of biodefense funding was strictly biodefense as opposed to multi-use.
- Excluding COVID-response funding, the FY 2023 President's budget requested about \$11 billion in discretionary biodefense funding, covering threat awareness, prevention, preparedness, surveillance and detection, response, recovery, bioforensics and attribution, and mitigation.
- The final 2023 budget provided about \$950 million for the Biomedical Advanced Research and Development Authority via the Public Health and Social Services Emergency Fund, much of which the report argues is not relevant to catastrophic risks.
- Total philanthropic funding for biosecurity, excluding direct response and HIV-related pandemic funding, remains in the tens of millions.

- The report carries 304-plus endnotes and an appendix with a grantmaker dilemma checklist, an interventions evaluation, and plain-text R code for its charts.

What the author is unsure about

Ruhl is unusually transparent about the limits of his own evidence. He says the politics of government budgeting mean line items do not always mean what they appear to, making a rigorous analysis of health security spending difficult, and quotes a long-running biodefense funding analysis noting that programmes have been combined with other line items and are no longer distinguishable, leading to the appearance of higher funding. He asks how much apparent post-COVID biosecurity funding is truly biosecurity funding and answers that the question is difficult to answer definitively. He says his own funding spreadsheet is incomplete and, because it came from manual searches, likely biased, and that advisory-service grants made by advised individual philanthropists may not appear in databases that mostly capture foundations. He notes that large uncertainties surround the robustness of medical countermeasures to worst-case threat scenarios. He states that the report deliberately avoids discussing specific biological threats or blueprints for harm, discusses risky pathogens only in abstract and general terms, does not disclose or draw attention to specific

biodefense vulnerabilities, uses well-known historical examples where possible, and asked pre-publication reviewers to watch for hazardous information, so some of the underlying detail is withheld by design. And his conclusion's portrait of a safer future is offered as something to imagine, not forecast.

For the discussion

- A point worth making: the most actionable line in the report is a measurement, not a value judgement. Between 2001 and 2014 only about 21% of what was counted as US biodefense funding was strictly biodefense, which means the headline numbers in this field routinely overstate what is actually being spent on the problem.
- A tension worth raising: Ruhl asks funders to prioritize interventions with near-term positive externalities that can win broad political support, and also to focus on worst-case scenarios. Those pull in opposite directions, and germicidal light is one of the few things that satisfies both.
- A question worth asking: the report's own practice is to avoid naming specific threats or vulnerabilities, and its grantmaker checklist asks whether a grant could be misconstrued as supporting offensive activity. How much of the field's difficulty in attracting funding comes from

the fact that the most persuasive evidence is the evidence it is not willing to publish?

1.6-3 · CORE · COURSE TIME 90 MIN

Securing Civilisation Against Catastrophic Pandemics

Anjali Gopal, William Bradshaw, Vaishnav Sunil and Kevin M. Esvelt · 2023

[Read the original paper →](#) · the course assigns: whole report (Geneva Paper 31/23), from the executive summary, introduction, likelihood sections, defences and conclusion, plus the appendices' estimation tables

The same report that 1.3 #8 excerpts, read whole: it argues that the pandemics capable of ending civilisation are almost certainly not natural, that the likeliest sources are an accident in a weapons programme or a deliberate release once blueprints are published, and that nations can defend against both for a tiny fraction of their defence budgets.

Average summary

This is the same Geneva Paper as 1.3 #8, where the course assigns only pages 10 to 12. Here it is optional, so this summary covers the rest of the report. The two scenarios in one line each: a "Wildfire" pandemic is so lethal and contagious that essential

workers stop working and essential services fail, and a "Stealth" pandemic spreads for years with few symptoms before causing mass death or disability.

The key takeaways are blunt: nations cannot yet contain natural, accidental or deliberate pandemics, access to severe pandemics will expand with the ability to program biology, and if too many essential workers die or refuse to work, societies will collapse. Nations failed to halt SARS-CoV-2, an agent with an estimated **R0 of about 2.79** introduced in one city, and many viruses are far more lethal: variola major killed approximately 30% of patients at an estimated R0 of 3.5 to 6, and caused 500 million deaths in the century before its eradication. Figure 1 makes the timing problem concrete: vaccines for SARS-CoV-2 reached a billion people in 586 days, the US National Biodefense Strategy's "moonshot" calls for 130 days, and Omicron infected a quarter of North Americans and half of Europeans within 100 days of detection.

The likelihood sections are the report's most distinctive contribution. It defines a Wildfire pathogen as one with a case fatality rate above 20% and a basic reproduction number exceeding 5.5, benchmarked on Omicron's estimated R0 of roughly 4 to 5.5 with an upper bound of 6.8. Only one pandemic pathogen in four thousand years of dense cities and trade routes, smallpox, has been both highly lethal and sufficiently contagious, so the annual chance of a

natural Wildfire pandemic sits between negligible and a theoretical maximum of 0.1%. Accidents in a weapons programme working with such an agent yield annual likelihoods mostly in single-digit percentages, and once genomic blueprints are public while unscreened DNA synthesis remains accessible, a deliberate release may run from the low single to low double digits per year. For Stealth there are no historical examples, so natural origin carries only a very low upper bound, under 0.01% per year, against 0.28% to 13.4% for a deliberate release.

The defences section is the practical half, and where the report's vocabulary comes from. Against Wildfire: identify primary essential workers (food, water, power, law enforcement), secondary workers who maintain their equipment and supplies, and life-saving workers whose absence kills people but does not collapse society, then supply them with pandemic-proof personal protective equipment, or P4E, at the onset. Against Stealth: deep metagenomic sequencing of wastewater from aircraft and travel hubs, looking for any nucleic acid fragment growing quasi-exponentially; a pre-recruited network of expert responders; fast medical countermeasures; and germicidal lights and ventilation in public buildings. The conclusion states the thesis in miniature: civilisation-threatening pandemics are extremely unlikely to emerge from nature, but accidental or deliberate release is disturbingly likely, and nations

can defend against both for a tiny fraction of their annual defence budgets.

Why it's assigned

On the 1.6 page it is offered as a long overview for anyone who wants the whole picture, and it is the one optional reading that doubles as a map of the course. Its four defensive moves are Units 2 to 5 in order, and its two scenarios are the worst cases those units are measured against. Reading the whole report rather than the assigned three pages gets her the numbers behind the scenarios, the full essential-worker taxonomy, and the germicidal light evidence she will meet again in Unit 4.

Words to know

- **R0 (basic reproduction number)** — the number of people a typical infected person infects in a population with no immunity. The report uses it to set its Wildfire threshold, and carefully distinguishes it from R_e , the effective reproduction number observed while precautions are in place.
- **P4E (pandemic-proof personal protective equipment)** — protection good enough that an essential worker can do their job with negligible risk of infection, with six core requirements including no fit testing for over 90% of users, better than 99.97% infectious particle removal,

and cost parity with N95 masks on an annualised basis.

- **Primary, secondary and life-saving workers** — the report's taxonomy: primary workers deliver food, water, power and law enforcement; secondary workers maintain and repair the equipment and supplies those depend on; life-saving workers, such as healthcare and social care, save lives without whom society would not actually collapse.
- **Nucleic acid observatory** — a monitoring system that records every biological sequence showing quasi-exponential growth, rather than searching for known threats. It is the report's answer to the detection problem, because any pandemic-capable agent must grow that way.
- **Expert responders** — a pre-assembled network of influential scientists and physicians who can assess a detected threat and testify within days. The report treats credibility as infrastructure, to be built before it is needed.
- **Germicidal light** — ultraviolet light used to inactivate airborne pathogens, either upper-room 254nm or 265nm LEDs above head height, or far-UVC below about 235nm that proteins absorb before it reaches human DNA.
- **Fast grants** — emergency research funding released quickly to pursue every promising therapeutic approach. The report pairs them with hope as a tool of social order.

- **Positive pressure (in PPE)** — pumping sterile air into an enclosed volume around the wearer's face so infectious particles cannot reach the mucous membranes, as opposed to relying on a seal. The report argues most adequate designs will need this, because fit testing is not feasible at population scale.
- **Dead-hand switch** — an automatic retaliation system that triggers without human decision. The report raises the possibility of a rogue state building one to cause a Wildfire pandemic, which could then leak by accident.

Numbers in this reading

- SARS-CoV-2's initial estimated R_0 : about 2.79, introduced in one city. By the end of 2022 it had infected most of humankind and directly or indirectly killed perhaps 20 million people, an effective global fatality rate of 0.25% including reinfections.
- Variola major: approximately 30% case fatality, estimated R_0 of 3.5 to 6, responsible for 500 million deaths in the century before eradication. The 1971 Aralsk outbreak, involving a potentially weaponised strain, was transmitted by vaccinated individuals and showed 100% lethality in the three unvaccinated patients.
- Rabbit haemorrhagic disease virus: highly contagious, transmitted asymptotically in young

animals, 80 to 100% lethal in adults. The report's point is that worse pathogens already exist in other species.

- Figure 1 timeline: vaccines developed and delivered to a billion people in 586 days; the US National Biodefense Strategy's target is 130 days from sequencing; Omicron infected a quarter of North Americans and half of Europeans within 100 days of detection.
- Wildfire definition: case fatality rate above 20% and basic reproduction number exceeding 5.5. Omicron's R_0 estimated at roughly 4 to 5.5, probably no more than 6.8, from R_e estimates of about 3.13 to 3.4 after zero-COVID controls were removed and 1.57 before.
- Annual likelihood of a natural Wildfire pandemic: between negligible and a theoretical maximum of 0.1%, since only one pathogen in four thousand years qualifies.
- Table A2, Stealth likelihoods per year: natural spillover tiny to under 0.01%; bioweapons lab accident 0.12% to 4.8% per lab per year; bioweapons programme accident 0.28% to 5.2% per programme per year; deliberate release once credible blueprints are public 0.28% to 13.4%.
- Collapse consequence: losing the essential workers who generate and distribute industrially produced fertiliser alone would cause 3 billion people to starve.

- P4E requirements: ready to wear with no fit testing or training for more than 90% of users; better than 99.97% infectious particle removal; wearable for long hours every day for months; cost parity with N95 masks on an annualised basis at around US\$200 each; meeting NIOSH requirements for positive pressure devices; and an unobstructed field of view with the full face visible.
- Why masks are not enough: HEPA-filtered air reduces infection risk by nearly 10,000-fold, while elastomeric respirators cannot offer most users even a tenfold reduction because they require fit testing, which often fails and which is not feasible at scale; tightly sealing masks cause adverse skin reactions and headaches in over 90% of users who wear them for extended periods.
- Workers who never share space with another person, or enter rooms vacated less than four to six hours earlier, should not require P4E.
- Fewer than 200,000 biomedical scientists in the United States, so protecting the research enterprise needs far fewer P4E units than protecting all primary essential workers.
- Germicidal light: in the 1940s, 254nm lamps installed above head height in high-ceilinged schools helped suppress tuberculosis and possibly chickenpox, a pathogen with an estimated R_0 of about 7, but suppression was much less effective when children travelled in buses without them.

Current North American workplace-safe thresholds of 222nm light can inactivate 90% of aerosolised viruses within a minute, roughly nine times as effective as aircraft ventilation systems that achieve one complete air exchange every three minutes. Figure 7's estimates at 23 mJ/cm² give about 95% of surface-borne, about 80% of aerosol and about 0% of droplet pathogens eliminated; at 478 mJ/cm², over 99%, over 99% and about 5%.

- US employers suffer an estimated US\$300 billion in productivity losses to infectious disease each year, which the report uses as the commercial incentive for installing protective lights.
- Risk-mapping app concept: a person's risk updates for everyone within five connections of a reported infection.

What the author is unsure about

The report's uncertainty is structural rather than incidental, and it says so. Its likelihood estimates are presented as wide ranges with explicit bases, and for Stealth it states that without any historical examples only an upper bound can be established for a natural origin. It says it does not know what degree of lethality is required to intimidate enough essential workers to stay home to cause collapse, and Box 1 turns that into four open research questions, including the relationship between fatality rate and willingness to work and how social media will affect it. It notes

Omicron's R_0 is hotly contested because immune evasion inflates naive transmissibility calculations. It says it is unclear how people would react once a Stealth agent's effects appeared. On communication it concedes that even the best messaging campaigns in high-trust populations will not persuade everyone, and that the challenge will be to save as many lives as possible whether or not people believe a problem exists. On far-UVC it says preliminary experiments suggest higher levels may prove safe and asks nations to fund the research to find out, inviting and addressing criticism from experts and the public. Its own prescriptions rest on things that do not exist yet: regulators currently assess device function by quantifying filtration efficacy and will need to develop alternative metrics to evaluate sterilisation. And the saved text contains a visible editorial note where a citation was still outstanding, which is worth knowing if she quotes the passage about masks and Omicron.

For the discussion

- A point worth making: this report's most striking number is not about a pathogen. Losing the workers who produce and distribute industrial fertiliser would cause 3 billion people to starve, which is why the authors treat supply chains, not hospitals, as the thing a pandemic defence protects.

- A tension worth raising: the authors argue that credibility has to be built in advance, with expert responders recruited and institutions briefed before anything happens, because the one thing that cannot be manufactured during a Stealth pandemic is public trust. That makes communications infrastructure as much a defence as a ventilation system, and it is the one no agency owns.
- A question worth asking: their estimates put a deliberate release once genomic blueprints are published at up to the low double digits per year, far above any natural scenario. If publication is the variable that moves the risk most, what does that mean for a research culture whose norm is to publish sequences and protocols in full?