“The science that celebrates itself”: the emergence of Biosecurity and Artificial Intelligence as a convergent public policy issue

Os Keyes, Samuel Angeli-Nichols and Nicholas Evans

University of Massachusetts, Lowell

BioSocieties (in press)

“The science that celebrates itself”: the emergence of Biosecurity and Artificial Intelligence as a convergent public policy issue

Abstract: This paper provides a multiple-methods analysis of an emerging domain of biosafety and bioscience scholarship – the intersection of biosecurity and Artificial Intelligence (AI), or BSxAI. Conducting both funding and genealogical analyses of the literature, we explore how this intersection came to be problematised, and the intellectual and financial sources mobilised in popularising the space. We argue that rather than constituting a conventional scientific domain, BSxAI is an area intentionally cultivated by a narrow range of funding bodies with particular ideological interests tied to the Effective Altruist and longtermist movements. The result, as we discuss, is a concerning degree of intellectual and structural brittleness – one with severe possible consequences for the life sciences, and AI policy, more broadly.

Keywords: biosecurity, artificial intelligence, effective altruism

Introduction

In November 2024, the AI Safety Institutes of (respectively) the United States and United Kingdom published a joint report evaluating the Claude 3.5 Sonnet machine learning model developed by the company Anthropic. Surprisingly, the very first capability on which the Claude model was evaluated related to the model’s “Biological capabilities”: its potential, in the eyes of the evaluators, to aid in biomedical innovation, together with its risk of “being misused to help harmful actors to synthesize and use potentially dangerous biological agents” (2024: 5).

This is hardly the only government action around the potential use of AI to create biological weapons. Among other instances, in September of 2023 a G7 meeting in Hiroshima reported

that “the threat to biosecurity as one of the most important priorities in the field of generative AI,” and that “several G7 members stressed the importance of preventing the use of generative AI to create chemical or biological threats (e.g. viruses)” prior to claims about cybersecurity, misinformation, or social stability (G7 2023). The US National Security Commission for Emerging Biotechnology released 4 white papers on so-called “AIxBio” in 2024 outlining risks, benefits, and policy priorities for addressing the security implications of frontier AI models and biological design tools (NSCEB 2024); the executive summary of that the Board’s final report would explicitly invoke the nuclear age as analogous to modern biotechnology, claiming that the United States “finds itself competing with a rival [today, the People’s Republic of China] over a new form of engineering that will create tremendous wealth, but, in the wrong hands, could be used to develop powerful weapons” (NSCEB 2025: 7). Section 236 of the FY2025 US National Defense Authorization Act, released in May 2024, required the Secretary of Defense to develop a pilot program on development of near-term use cases and demonstration of artificial intelligence toward biotechnology applications for national security. And a bipartisan bill introduced in the US Senate in 2025 required the US Secretary for Health and Human Services to develop a strategy for public health preparedness and response to address the risks of misuse of artificial intelligence (AI), particularly with regard to the development of biological or chemical weapons—again, prioritizing biological weapons threats from AI over other risks ostensibly under the purview of that agency, such as the misuse of sensitive health information through AI *(*U.S. Senate 2025). As with AISI, these policy interests are not limited to the United States; the United Kingdom Biological Security Strategy, too, emphasizes that the convergence of "bioscience with...machine learning and AI...[is] creating new...risk” (HMG 2023: 34). And in April 2026, the Australian Government announced the formation of a new taskforce explicitly to tackle “catastrophic” AI-enabled bioweapon (Smith 2026).

For scholars who have been studying AI – or the life sciences – this raises some obvious questions. Why is the prospective computational design of novel pathogens the first priority of the national bodies tasked with assessing AI safety in an environment where active cybersecurity vulnerabilities, deep fakes, and misinformation already pose real, material risks on individuals and communities, some of which constitute direct and immediate threats to national security (ODNI 2024: 31-33)? How did these risks come to be perceived by policymakers on multiple continents as problems worth tackling so urgently; a threat so profound that multiple bills in the US Congress, which has infamously been reticent to regulate biology (Wright 1986), seek to tackle it? How did the space of "AIxBio" come to exist?

A naive view of this might imagine that the risks of the biological implications of AI are, as a problem, simply "out there" to be discovered and studied by scientists. For scholars from Science and Technology Studies (STS), things are more complex: much research has demonstrated that scientific questions are not born, but made; problems not "out there", but actively rendered, or "problematised". The work of Donald MacKenzie (1993), for example, has demonstrated how even a concept as seemingly-obvious as "accuracy" is in fact the result of much active work by communities of practice – work that itself reconfigures the priorities, methods, and norms of those communities. The same has been demonstrated by many other scholars, studying a multitude of scientific spaces and problems. Problematisation has shown how intentional choices about which issues to elevate to the fore have affected fields as diverse as safety in automotive engineering (Johnson 2009); diagnosis in psychology and psychiatry (Rose 1998); and priority setting in global health research and practice (Lakoff 2017).

One subset of the problematisation literature that has widely been applied to biosecurity concerns *securitization,* the mechanism through which phenomena are labelled "security

threats" (Buzan *et al* 1998). Considerable work has been pursued on how biosecurity professionals "discover" threats to national security arising from the life sciences (Rappert 2007; Lentzos 2009). Key trends in the securitization of biology include the role of framing cases in grounding intuitions about the possibility of biological weapons use (Annas 2010); claims about the distribution of technological capabilities in society as a way to generate security claims (Vogel 2012); and the relations between larger political trends such as the commercialization of synthetic biology on norms in global security and interstate politics (Edwards 2019). However, to date there has been relatively little demonstration of the fine-grained *mechanisms* by which problematization occurs in biosecurity, with those entries that make such attempt typically focusing on technical, rather than policy and political mechanisms (e.g. Ouagrham-Gormley 2014). Moreover, the emergence of AIxBio into policymaking spaces is of interest for the way it has appeared separately to extant biosecurity conversations, including laboratory safety and personnel reliability debates, and international non-proliferation, from which it has appeared separately.

Drawing on the scholarship on problematization, we use this paper to descriptively answer the question of how "AIxBio" came to be prioritized in AI and bioscience policymaking. We use a multiple-methods approach consisting of an analysis of funding trends that inform the emergence of this corpus of literature, and a conceptual genealogy of the author and institutional network that underpins the space. We use these to describe the emergence of the phenomenon of "AIxBio" (which, given its predominant focus on health risks from AI, we have renamed Biosecurity x Artificial Intelligence or BSxAI) from its early developments to its current status as a critical policy issue. We conclude that this emergence can be explained by a concentrated application of scholarship motivated by a small cluster of elite philanthropic funders affiliated with the longtermist and effective altruist (EA) movements.

We argue the majority of the BSxAI literature, extending from its earliest papers to current policy initiatives, is grounded in a series of common assertions regarding "Global Catastrophic Biological Risks," a term coined and promoted almost exclusively through think tanks and academic research centres who receive extensive funding from private Effective Altruist philanthropic organisations (Schoch-Spana *et al* 2017). Moreover, the story of the BSxAI corpus is dominated by its alignment with other actors in the ideological movements of Effective Altruism, longtermism, and (to a lesser degree) rationalism. These three movements are the most public components of the TESCREAL bundle of ideologies (described more below) whose politics have recently received sustained criticism for their tendency to prioritize utopian and apocalyptic visions of technological development over central concerns and projects of real, often marginalized communities (Gebru and Torres 2024), extend and reproduced of social inequalities, and connect to racist and eugenicist movements (Wilson and Winston 2024; Anthony 2024).

Our paper proceeds as follows. First, we describe the project beginning with collection of a corpus of documents that focus on BSxAI from 2020, the earliest such publication, through to the end of 2024, the end of our study period. We identify a cluster of disconnected authorship communities that, on a deeper funding analysis, are connected through their relationships to and financial support from EA and other ideologically aligned philanthropies. We then conduct a genealogy of BSxAI from its prehistory through to current day, interrogating the evidence base and ideologies at play. To conclude, we discuss both the circularity of that evidence and the way that BSxAI intellectually and practically launders a particular bundle of ideologies – the "TESCREAL bundle" - into the biosecurity literature. Such laundering, we argue, has wider implications for how we understand not only the intellectual viability of BSxAI, but the ways that scientific research is funded, interrogated, and integrated into policymaking.

An emergent corpus

Starting from the provocation in our introduction – that there has been a sudden uptake in policymaking interest in this domain – we began our work by seeking to understand whether this rapid increase in attention appeared similarly in academic work in the area. We collected a corpus of publications concerned with the intersection of AI and biosecurity, with a focus on papers concerned with the use of AI to enable the creation of biological weapons. We used two structured searches conducted through the National Library of Medicine/PubMed and the Dimensions.ai platform, combined with a broad grey literature search by two of the authors. Search terms focused on "biosecurity" and "artificial intelligence," and related terms (see Supplement for full strategies). After deduplication, we extracted 135 publications for review. We coded publication type, peer review status, authors, acknowledgements, editors, and declared funding sources as base data; final inclusion was based on a paper whose primary focus was AI and biosecurity (n=74). One paper was later removed from our sample during analysis when the team realized it concerned chemical security and chemical weapons development as it relates to artificial intelligence, rather than biological security, providing a final corpus of 73 articles. As the research project continued, we also made (and draw on the results of) a Public Records Act request to the University of Washington, home of both some of the researchers identified in our corpus, and one of the major workshops on the topic in recent years.

A flurry of interest

Our corpus revealed several interesting findings, the first and most immediate of which was the speed at which BSxAI emerged. From 2020-2022, only 6/73 (9%) of documents in our final analysis were published; in 2023, 28 (38%) papers were published, followed by 39 (53%) of documents in 2024. A preliminary analysis of 2025 publications found 57 additional publications – a 50% growth rate - that will be included in future analysis. The field is thus, at

least on its face, exploding, generating more documentation in each successive year than in all previous years over the study period (see Figure 1).

We found this unusual relative to larger trends in biosecurity, which is often considered a niche policy issue over which only a small number of researchers concern themselves (Lentzos 2019). Biosecurity, moreover, is typically not well funded, and suffers from a panic-and-neglect cycle where controversies – for example, the Aum Shinrikyo sarin attack (Leitenberg 1999) fears of the spread of biological weapons technology and expertise following the collapse of the Soviet Union and the shuttering of its Biopreparat program (Leitenberg and Zilinskas 2012), and the anthrax attacks in the United States in 2001 (Enemark 2017) - generate funding that fails to produce continuing support for the field (e.g. Dzau 2026). Interest and investment have been the result of particular "paradigm cases" or moments of perceived urgency, many of which have been well described elsewhere (Evans 2016; Koblentz 2017).

But there is no such equivalent triggering incident in the BSxAI literature. In our review of the literature, only one tangible case of what might be considered a framing event emerged—Bloomberg News reporting on a White House meeting in which a security analyst showed how he could use AI to develop a biological pathogen—but, surprisingly, this example was *not* present in the actual literature in our BSxAI analysis. Why, then, the sudden flurry of interest? The answer, we content, is in part a matter of *funding* driving popular interest, rather than the other way around.

Funding and ideology

We noticed some initial unusual patterns in formal funding disclosures in the BSxAI corpus. Only 17 (23%) of documents in our corpus reported funding sources. Only a minority (22%) of the documents in our corpus were peer reviewed, and while Committee on Publication Ethics standards require journals to solicit funding declarations for research articles, not all

sections of the masthead report these in the same way. Preprints (16%) and white papers among other grey literature sources (36%), finally, often have no requirements on them for funding declarations, which may have explained the low number of funding declarations.

But what *is* unusual is where this funding came from: 70% of works reported direct support from an organization called the Open Philanthropy Project (OPP), who recently announced their transition to the name Coefficient Giving (Walsh 2025). OPP appeared four times as often as any other funding source. And it is not alone as an unconventional, nongovernment funding source in our corpus. BSxAI sits at the intersection of the biological sciences and artificial intelligence, both spaces that are highly prioritized by conventional governmental funding mechanisms (e.g. Long and Marzi 2021). But of the papers with funding disclosures, only four reported any funding from such sources – and those that did also reported funding contributions from less common sources (see Table 1)

OPP and the majority of other funders in the corpus are ideologically aligned philanthropic funders. OPP itself is often identified as a central funding mechanism of the Effective Altruism (EA) movement (Gleiberman 2023; Syme 2019). Building on the philosophical work of William MacAskill, Toby Ord, and Peter Singer, amongst others, the EA movement advocates (as the name suggests) "using evidence and reason to figure out how to benefit others as much as possible, and taking action on that basis" (MacAskill 2017: 2). Adherents have attracted extensive funding, predominantly from the highly wealthy, including large-scale funding from billionaires – such as those who founded and oversee the OPP itself, which began as Good Ventures funded by Dustin Moskovitz, one of the founders of Facebook (Matthews 2015).

EA is often associated with a range of like-minded movements that Gebru and Torres (2024) have labelled the "transhumanism, singularitarianism, (modern) cosmism, Rationalism,

Effective Altruism, and longtermism" (TESCREAL) bundle. MacAskill, who is credited as the founder of the EA movement through organizations such as *Giving What We Can* and *80,000 hours*, was also one of the earliest authors who explicitly identified as longtermist (Greaves and McAskill 2021). Those same organizations have been grouped with OPP and its predecessor organizations, and with the rationalist forum Less Wrong, the singularitarian and transhumanist Singularity Institute (Boenig-Liptsin and Hulburt 2016), and utilitarian message board Felicifia as part of "the convergence from 2008 to 2012 [of] at least 4 distinct but overlapping proto-EA communities," by Jacy Reese Anthis (2022).

Movements within TESCREAL do not precisely overlap, but tend to share a strong focus on technological risk and technical solutions to social problems; a strain of utilitarianism that seeks to perfect risk-neutral, probabilistic, expected value analysis as a form of moral thinking; prioritizing attention to moral problems that its adherents consider extreme, urgent, and neglected by others; and a view of morality that places the welfare of sentient beings that will exist in the long term future (thousands to billions of years) on similar standing to currently existing humans. The movement is also, as Gebru and Torres note, known for its strong ties to twentieth century scientific and philosophical movements that advocated for eugenics: transhumanism, for example, was popularized by eugenicists such as Julian Huxley (Gebru and Torres 2024: 5). These criticisms dovetailed with public controversy in 2024 when the Future of Humanity Institute at Oxford—founded by Nick Bostrom and known for producing high volumes of EA, longtermist, and transhumanist thinkers—was closed, in part because of accusations of eugenicist sympathies (Anthony 2024); and broader concern by scholars on the motivations of the movements and their ultimate goals (Wenar 2024a).

Prompted by the presence of OPP, we sought to understand the ideological background of other funders, and whether they are articulated (or articulate themselves) as part of the TESCREAL bundle of ideologies. Alignment was established through organizational

affiliation: the acronym describes a set of philosophical and political positions of which many proponents simply self-identify in survey research (MacAskill 2019), and central organizations that have longstanding historical relationships to the movement, documented extensively, also by self-described members of those movements (Anthis 2022). We thus were able to identify organizations as TESCREAL-aligned to the degree to which they claimed alignment with one of the movements described by the umbrella terms in their official documentation, websites, or in statements by key figures at those organizations.

We found that the overwhelming majority of other funders identified in Table 1 also define themselves in relation to EA or other TESCREAL movements, particularly Effective Giving (Ergo Giving as of late 2024 (Nash 2024)); and what we might term "second generation funders," Horizon Institute, ML Alignment Theory Scholars Program, Long-Term Future Fund, and Sentinel Bio, all of whom received OPP funding in addition to disbursing their own along the lines of TESCREAL priorities. The predominance of these funders in our sample gave us reason to dig deeper into the relations between such organisations and the BSxAI literature.

Formal funding disclosures, of course, are highly limited. Different researchers can have different ideas of what it is appropriate to disclose, and disclosures only track direct relationships, from granter to grantee, and are of limited use in tracking the broader context and spaces in which authors and funders operate. Moreover, with peer-reviewed journal articles the minority of publications in our corpus, we were limited in direct analysis. To try and dig deeper, we mapped that broader context: the ways in which funders and researchers are entangled and relate before, after and beyond directly reported funding.

We drew on two sources in aid of this goal: OPP itself, which historically published aggregated metadata on the grants it funded, and – following the apparent thread of

connection to TESCREAL – similar data from the Survival and Flourishing Fund (SFF), a longtermist funder that has worked in overlapping spaces. SFF as a longtermist institution is particularly attractive, because EA and longtermism have similar origins, with a seminal paper on longtermism written by Greaves and MacAskill (2021) funded through the Global Priorities Institute, an EA-focused think thank that received >$6 million USD in funding from OPP from 2020-2025.

To track how this funding connected to authors in our corpus, we used LinkedIn and similar tools to trace the affiliations of authors within our datasets over time – where they had worked and studied. Linking the two together on location resulted in a dataset of 340 authors, editors, and collaborators, across 190 separate institutions over time, against which we cross referenced 986 OPP and SFF grants (See Online Supplement). This provided a broad map of the authors, editors, and acknowledged intellectual influences in the BSxAI corpus, and their relationship with two particularly wealthy TESCREAL-aligned funders.

Our data demonstrated –beyond the 17 papers that explicitly declare funding that acknowledged a TESCREAL-aligned funding course– significant penetration of OPP and SFF in the BSxAI corpus. Taking first and senior (last) authorship as key markers of contribution to a paper and the project from which is derives, a total of 64% (n=47) of *all* BSxAI publications within our corpus had first or senior authors whose employers had, before, during or after work on the publication in question, been funded by OPP and/or SFF. Moreover, 71% (n=52) of all publications in the corpus had at least one author who was, is, or has been funded by OPP and/or SFF.

We suspect that our results are an underestimate given that philanthropies can fund institutions or investigators at their discretion, unlike federal grants that typically name a PI and other senior personnel. As an example, a 2024 editorial in *Science* on BSxAI included

authorship from leadership of the University of Washington's Institute for Protein Design (Baker and Church 2024). While the authors themselves were not reported in grant databases as receiving funding, the Institute itself received $11.3m from OPP in 2018, a further $3.2m in 2023, and a series of smaller (i.e. six figure or lower) grants from OPP in the years between. Yet this did not appear in our database due to the lack of explicit connection between author, institution, and grant identifiers. This also sets aside, of course, that the funding histories of the 13 other TESCREAL-aligned funders in the BSxAI corpus are not publicly traceable, and these do not represent the full gamut of possible TESCREAL-aligned funders in the space that may have provided general or other support, or funded work that did not require funding acknowledgement.

To summarise, then, a substantial majority of the literature *articulating BSxAI as a scientific, security, and governance problem* is funded, or directly connected to funding by, a small number of organisations. These are highly concentrated among philanthropic or private groups with an EA or longtermist ideology, the most prominent of which is OPP. While further ties to other TESCREAL funders and movements are beyond the scope of this paper, the overwhelming majority of papers' identification with two of the best funded movements in the cluster gives us reason to think that the others are likely represented.

Recognizing the connection between this field of inquiry and a small group of affluent, ideologically-aligned funders requires us to ask what about this space makes it attractive to these funders in particular, and whether there are deeper connections between the ideas of technology and risk that motivate interest in BSxAI, the priorities of these funders, and the adherents of the movements they champion. Answering this is the focus of the next section.

A Genealogy of BSxAI as an emerging issue

In principle, there is nothing objectionable to using evidence in prioritising charitable action. Scholars such as Lief Wenar (2024b) will often concede that this component of the EA movement acts as a vital concept for new members—why on earth would we not want our charity dollars to actually *do something?* Critics of the EA movement, however, highlight two major issues with the movement, and how adherents articulate and categorise evidence from which to set priorities. The first is that EA adherents tend to rely exclusively on large-scale, quantitative data about what problems – and solutions – to prioritise and fund. In a range of areas, from animal welfare to public health, this has led to a deprioritisation of the perspective and expertise of those "on the ground", directly confronting and dealing with the injustices in question (Adams *et al* 2023).

This same prioritisation of seemingly unambiguous and quantitative data also leads to what critics see as an over-focus by EA adherents on more speculative causes, from asteroid strikes to "runaway" AI or nanotechnology, over more immediate questions of poverty, health inequality or racial injustice. The hypothetical nature of these speculative issues, combined with their theoretical scale of encompassing the entirety of humanity—not just today, but for all points in the future—makes them paradoxically far more tractable to an EA worldview, since hypothetical scenarios are infinitely easier to model in a mechanistic, and seemingly-certain, fashion. Put another way, EAs have a recent track record of prioritizing risks based on quantitative estimates of the impact of these speculative scenarios, which due to their alleged "existential" risk, put a thumb on the scale against which any other kind of issue pales in comparison (Wenar 2024a). As a 2015 report on a global EA conference noted, while EAs initially prioritized addressing global poverty, the threats of artificial general intelligence make global poverty a "rounding error" to EAs in terms of the potential risks of the former,

even though the latter is ongoing and harmful to real, rather than hypothetical future, people (Matthews 2015).

Given the significant financial influence of these movements on the BSxAI corpus, we conducted a genealogical analysis of the literature in the space to better identify the relationships between the ideologies of those funders and the work produced with their support. Genealogy in this sense entails qualitative analysis of the concepts and discursive frames articulated in publications: what their authors prioritise and problematise, and what concepts (and where) they draw from in doing so. Such an approach is commonly used in Science and Technology Studies to understand the emergence, and shaping, of an issue (Waidzunas and Epstein 2015; Keyes 2020).

Starting from the sample used in our network and funding analysis, we traced out the major themes in each piece, the development cycle of the ideas behind BSxAI, and the sources outside the space that researchers drew on to ground and justify their research. We ultimately found, in summary, that the overlap between TESCREAL ideologies and BSxAI is not a late-developing coincidence. Much of the justification of BSxAI, rather, as a set of problems to be solved originates within TESCREAL literature and those movements' conceptions of risk and futurity. This – and the way that it has led to BSxAI literature replicating common criticisms of these ideologies – raises serious concerns about the concrete applicability of these works.

Global Catastrophic Risk

In their introduction to *Contested Futures*, Nik Brown, Brian Rappert and Andrew Webster note three areas of "powerful commercial, policy-related and intellectual narratives associated with the anticipated future impact of three scientific fields: biotechnology, new materials, and ICTs" (Brown *et al* 2000: 11) Given this, it is perhaps unsurprising to see "biotechnology" and "ICTs" become hybridised into a single, urgently-phrased and future-

oriented narrative in the form of BSxAI (Attal-Juncqua *et al* 2026; Jeffrey *et al* 2023; Responsible AIxBiodesign 2023). What makes the BSxAI corpus particular, however, is its origins in discussions of a term that begins in TESCREAL spaces: "global catastrophic risks" (GCRs).

GCR, as a term, was popularised in Nick Bostrom and Milan Cirkovic's edited volume of the same name (Bostrom and Cirkovic 2008). They scope the term and volume as being concerned with "[risks] that might have the potential to inflict serious damage to human well-being on a global scale", including "pandemic infections, nuclear accidents...worldwide tyrannies, out-of-control scientific experiments [and] climate changes", (1) amongst others. Catastrophic risks arguably have their basis as an adjacent category to "existential risks," first formulated by Bostrom in 2002 as risks "where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential" (Bostrom 2002).

As a concept and a book, GCR is deeply grounded in longtermist, EA, and transhumanist thinking. Bostrom is a prominent figure in all these movements as are many of the book's contributors. On the longtermist front, Bostrom and Cirkovic emphasise their contributors assess risks "not only as they presently exist but also as they might develop over time", with time constituting a scale of not only decades but "even centuries" (5). Just as familiarly from our discussion of EA, the concept focuses on possible risks without particular attention to their likelihood, assigning urgency and relevancy to risks that "are almost certainly physically impossible" with the logic that "if we are not certain that there is no objective risk, then there is a risk at least in a subjective sense" (5). This focus (on long-term and improbable possibilities) leads to the contributors spending as much time on the possibility of nanotechnology consuming the world as on the rise of tyrannical and fascistic forms of government. While less explicit, transhumanism is invoked as a *solution* to the challenges of

GCR: a chapter by Yudkowsky (2008: 342) concludes, among other examples, with a meditation on AI as one of a number of paths for humans out of their evolutionary limitations, particularly with regards to aging and intelligence.

Although some of the catastrophic global risks Bostrom's colleagues highlight are "natural" ones, such as geologic events, many are the direct result of human action, with two areas of focus being high technology (such as nanomachines and Artificial Intelligence), and the biological domains technologies intervene in. In the case of AI, a rising narrative over the last decade has argued that the risks of AI are not the (many and well-documented) instances of AI systems exacerbating or depending on social injustices, but instead more apocalyptic and speculative. In particular, researchers in this space began increasingly treating the development of a super-intelligent Artificial General Intelligence (AGI) as inevitable. This concern of runaway AI focused efforts on resolving a feared 'alignment problem' - the worry that such a system, lacking the same values as humans, would be capable of causing immense harm and suffering. Such concerns have been explicitly framed through the lens of AI as a GCR (Turchin and Denkenberger 2020), but (even without this explicit label) originate in the same space – and with some of the same authors – as Bostrom's work (Yudkowsky 2016). These theories and theorists have influenced not only the literature around GCRs, but AI ethics, as frontier model companies—and OpenAI and Anthropic in particular, whose leadership are closely connected to TESCREAL movements—fixing the north star of frontier AI development, and thus AI ethics, as that of alignment. This trend is so pervasive that multiple sets of researchers in AI ethics have expressed concern about this approach distorting the field's focus on more concrete, and less speculative, harms (Gebru and Torres 2024; Westerstrand *et al* 2024).

In parallel, but distinctly, the biosciences became the subject of their own explicit sub-category of GCRs: "Global catastrophic biological risks" (GCBRs), defined in the paper of

the same name by a team out of the Johns Hopkins Center for Health Security (CHS) - which has received over $60 million from OPP as of 2025 (Schoch-Spana *et al* 2017) as:

> "those events in which biological agents—whether naturally emerging or reemerging, deliberately created and released, or laboratory engineered and escaped—could lead to sudden, extraordinary, widespread disaster beyond the collective capability of national and international governments and the private sector to control. If unchecked, GCBRs would lead to great suffering, loss of life, and sustained damage to national governments, international relationships, economies, societal stability, or global security" (323)

This definition is explicitly linked to GCRs by the authors through Bostrom and Cirkovik, a second paper by Bostrom (2013), and a GiveWell (recall, OPP's sister organization) blog post by Nick Beckstead (2015), an EA thinker who at the time worked for OPP's program on machine learning and is now is the CEO of the Secure AI Project (funded partly by OPP). Such a definition did not initially explicitly mention AI despite its connection to EA interests in nascent AI themes– although it did raise the spectre of "Novel or Artificial Organisms Harmful to Existing Life" (Schoch-Spana *et al* 2017: 327) as a type of GCBR that might emerge in the future. Instead, the GCBR definition papers largely addresses pandemic-like events, either naturally occurring or because of hypothesised 'lab leaks' or intentional (but conventional) bioweapons design programs. This primary focus on traditional biosecurity changed at the close of the 2010s, when AI and Biosecurity formally converged as areas of concern. A 2019 edited collection by members of the GCBR definitions team, for example, mention AI in the context of GCBRs across its chapters (see DiEuliis *et al* 2019), but do not interrogate the topic in favour of focusing on specific pathogens as GCBR threats under the rubric of more conventional themes in biosecurity. The following year, however, O'Brien and Nelson (2020) would explicitly link AI with Biosecurity, closely followed by two other

papers, written by Clarke and Whittlestone (2022) and Smith and Sandbrink (2022), each focused on the idea of a ”convergence" between the spaces.

All three papers are by authors with explicit ties to the EA movement: Nelson and O’Brien’s 2020 is authored by a member of the US Bipartisan Commission on Biodefense, which received $5.2 million in 2022 from OPP, and a scholar at the Oxford Future of Humanity Institute, founded by Nick Bostrom and the recipient of at least $19 million from OPP. Clarke and Whittlestone’s “A Survey of the Potential Long-term Impacts of AI” (2022), is authored by scholars affiliated with the Cambridge Centre for the Study of Existential Risk, the sister institute of the Future of Humanity Institute initially funded by Jaan Tallin, a self-professed EA (see Talinn 2018). Smith and Sandbrink’s “Biosecurity in an age of open science” (2022) was authored by a scientist who would hold contract positions at the OPP funded Centre for Long-Term Resilience and RAND Meselon Center, and would later work at the Mirror Life Dialogue Fund funded directly by OPP; and a researcher at FHI who would hold positions at the OPP funded Council on Strategic Risks Ending Bioweapons Program and spent time in the UK AI Security Institute, before becoming Entrepreneur-In-Residence at Sentinel Bio, a second-generation TESCREAL-aligned funder.

All three of these pieces explicitly reference catastrophic risk. O’Brien and Nelson, publishing their work in *Health Security,* the journal produced by CHS, are the most obvious in regard to GCBRs, citing the definitional paper when claiming that “With increased access to [AI and biotechnology], so comes the potential for a resulting high-consequence biological event, some of which may have such large consequences that they constitute a global catastrophic biological risk” (O’Brien and Nelson 2020: 220). Clarke and Whittlestone’s stated aim is to understand how AI could “Make a global catastrophe more or less likely (i.e. a catastrophe that poses serious damage to human well-being on a global scale, for example by enabling the discovery of a pathogen that kills hundreds of millions of people),” citing the

same Beckstead article that informed the GCBR definition. And Smith and Sandbrink's work centres thematically on scientific information in the face of "information hazards" that could cause harm through the misuse of biological knowledge. Information hazard as a concept is a term often credited to Bostrom (2011), and arguably best known through the rationalist online forum *LessWrong*, where it is explicitly used in the context of existential risks arising from AI (see Sandifer 2017).

AI meets Biosecurity – From Case Analysis to Abstract Theorizing

The risks of convergence in biology are not, however, an invention of EAs, the GCR literature, or the AIxBio movement. As described above, concerns about the radical harms posed by biological weapons emerged in the mid-1990s from several sources, including the Aum Shinrikyo doomsday cult, revelations of the Russian clandestine biological weapons program, and the anthrax attacks of 2001 in the United States.[i] Convergence subsequently arose as a topic after the beginning of the "dual use" concerns that have been central to the last 25 years of biosecurity policy, including numerous entries in our corpus (Soice *et al* 2023; Sandbrink 2023; Marijan 2023; Pannu *et al* 2024; Bloomfield *et al* 2024). In 2006, the National Academies of Science, Engineering, and Medicine report *Globalization, Biosecurity, and the Future of the Life Sciences* on dual-use research in the life sciences remarked that "other fields not traditionally viewed as biotechnologies such as materials science, information technology, and nanotechnology are converging with biotechnology in unforeseen ways and thereby enabling the development of previously unimaginable technological applications" (National Research Council 2006: 16).

Considerable resources, moreover, have been expended on characterizing the risks of biological molecules using computational systems (Rabinow and Bennett 2012; Tucker 2012; Garfinkel *et al* 2007; Evans and Selgelid 2015). These efforts were driven largely by

scientific experiments that demonstrated identifiable risks in virtue of their success. Perhaps most relevant for our purposes is the 2002 synthesis of the polio virus, a foundational experiment in synthetic biology and the first time scientists demonstrated that complete viral synthesis was possible without a wild template—opening the way for discussions of the possibility that state or nonstate actors could now generate existing biological pathogens, or novel engineered pathogens, through synthetic means (Selgelid and Weir 2010). Machine learning methods have been discussed in biology for decades (Ching *et al* 2018), as have concerns that with a sophisticated enough library of biological parts, computational methods could design pathogens more dangerous than their natural counterparts (Tucker 2010).

What sets BSxAI apart, then, is not the basic themes of biological risk and converging technologies; the power of computational methods in biology as a driver of risk; nor the extreme harms that could arise from the use of biotechnology, each of which has its own genealogy. Rather, it is its abstraction from a) the methods and focus of the historical biosecurity literature and b) concrete examples of risk. We should note that this observation is not an endorsement of previous modes of biosecurity *tout court*, which have received sustained criticism (e.g. Vogel 2012; Edwards 2019). Rather, this divergence, combined with our funding analysis, points at something interesting in the BSxAI corpus in need of explanation.

Consider O'Brien and Nelson's paper, the first to explicitly link AI and biosecurity in terms of the prospects of AI-designed biological weapons. As the title suggests, the authors discuss convergent risks that they portray as crucial for scholars in both fields to understand, with an emphasis on the need for "hybrid" and "convergent" approaches in doing so (O'Brien and Nelson 2020: 219). As O'Brien and Nelson frame it, these risks – arising from the combination of the increasing popularity and strength of AI technologies, and the application of these to the biosciences – can be exemplified in part by the possibility of AI enabling the

development of synthetic pathogens, by "lowering the technical barriers and tacit knowledge required" for pathogen development (223). The focus of the piece is not on specific risks, but instead on what the authors argue is a new approach to risk *assessment*, one they advocate for formal development of, to be pursued through a combination of US law enforcement and scientific bodies.

Without discussing the technical realities of AI (or the biosciences) as spaces of practice, we should examine how O'Brien and Nelson marshal support for their claim that the intersection of AI and biology is an active area of risk. Despite making the argument that pathogen synthesis is an urgent area of security concern (urgent enough to necessitate formal rulemaking and law enforcement investigation), the authors provide little evidence that concerning activities are occurring in this space. While they reference the *de novo* synthesis of 1918 influenza in 2005 and horsepox in 2016, both of which they refer to as "recent", they do so via citing a paper that is sceptical of those experiments as a cause for concern around the security of the life sciences (Koblentz 2017).

The primary resources O'Brien and Nelson point to secure their claims of *risk* are a book on the sociology of risk in the biosciences (Edwards 2019), and a paper published contemporaneously by Nelson herself that provides a *purely hypothetical* model of risk in the abstract (Sandberg and Nelson 2020). Neither contains any primary evidence that the events the authors worry about are tangible threats—that is, that AI provides *additional* reason to be concerned about extant biosecurity concerns—as neither work mentions AI at all. The former, moreover, is a work that is broadly critical of the speculative processes by which hypothetical future biological threats are constructed. In it, Edwards (2019: 8) argues that technology "is not only used as a tool to reproduce existing institutions and power relations; it becomes a site at which they are challenges, and this complicates quests to control." That

is, in a paper that posits a new framework for assessing threats, one of their primary points of evidence for threats is a work that calls into doubt the very systems they seek to create.

The same is true of the second peer-reviewed publication in this space, Clarke and Whittlestone's "A Survey of the Potential Long-term Impacts of AI" (2022). Pointing to a perceived risk-and-opportunity combination in how AI might impact scientific progress, they give as negative examples the possibility that "AI could speed up progress in biotechnology, making it easier to engineer or synthesize dangerous pathogens with relatively little expertise and readily available materials" (194) But this is a hypothetical – the only citations are to the work by O'Brien and Nelson who (as discussed) have no concrete cases of AI and biotechnological risk to provide, and a seemingly-unpublished paper that uses the GCR framework to discuss pandemic threats. Where O'Brien and Nelson couch their research in terms of risk assessment and health security that are broader than the narrow concerns of the TESCREAL bundle, Clarke and Whittlestone are more explicit about their ideological attachment. They emphasise that they are "particularly concerned with impacts of AI on the far future", and link that into concerns about "global catastrophe" (192) - concerns that they cite to the aforementioned blog post by Beckstead on GiveWell, the predecessor and one-time parent organisation to OPP (Matthews 2018).

It is worth noting that the point here is not *merely* that the BSxAI corpus engages in hypothetical reasoning—security scholarship and policy is frequently, and perhaps necessarily grounded in imagining the future (Loader and Walker 2007; Annas 2010). Rather, what is novel here is that the description of these threats is almost totally abstracted from a concrete example of risk. The assertion, rather, is that increasing sophistication of AI models *necessarily* constitutes a dual-use risk in biology, without any need for evidence, and thus requires governance. This, they claim, overcomes the need for specific threat assessments or "red-teaming" exercises; the hypothetical possibility should be enough (Götting *et al* 2025).

This emphasis on speculation rather than (or instead of) any evidencing dovetails with the ignoring of the larger historical trends in biosecurity. As an example, we can look at the two central concerns of O'Brien and Nelson and many other papers in our corpus; *deskilling* and *democratization*. The former is where technical barriers to *tasks* that might enable malicious uses of biology are lowered; the latter where the barriers to entry into the *field* are lowered. Both, many authors here argue, occur with BSxAI, resulting in a proliferation of risk. In some entries in our corpus, the logic of this model is made explicit (e.g. Soice *et al* 2023) and the conditions are assumed; in others, the logic is held to be sound, though when or if the conditions will obtain is held to be uncertain (Götting *et al* 2025).

But both deskilling and democratization are well studied phenomenon in biosecurity (Evans and Selgelid 2015), with democratization arguably foundational to the field (National Research Council 2006), and deskilling something that has been of concern since the advent of the synthetic biology movement (Garfinkel *et al* 2007). The question unanswered for BSxAI, and not addressed in the corpus thus far, is what exactly is novel about AI, beyond other methods of biological design, and to what degree does the current state of the art reflect a trajectory of significant change in these properties beyond what has been claimed to be an increasingly powerful biotechnology landscape for three decades?

AI, Biosecurity and Policy

Thus far, we have seen both the origins of "BSxAI" concerns and their percolation through academic and think-tank publishing. But it does not sufficiently answer the original question: how has it captured policymakers' attention so quickly and effectively? If anything, our descriptions of how uncertain, and weakly grounded, the literature in this space is, seems to make such capture even more improbable.

To understand how BSxAI comes to appear so prominently in formal state policymaking, we need to see this in terms of both framing and practice. By *framing*, we mean asking what it is about the rhetoric or claims of BSxAI that make them legible and attractive to state actors and role holders within state institutions (i.e. political appointees, members of parliament and/or congress, and policymakers). One obvious answer is that the framing of BSxAI follows from a larger history of securitization—itself a framing tactic whereby an issue is staged as an existential threat in order to generate endorsement of extraordinary treatment of that issue (Buzan *et al* 1998)—of both biology *and* computer science. "Health security" has a long history of claims that naturally emerging infectious diseases such as Ebola virus disease (Evans 2016) and scientific innovations such as viral synthesis (Selgelid and Weir 2010) are national security threats, in an effort to generate policy attention and action. Likewise, computer science has a long history of treating chipsets as strategic technologies that require export control (Daniels and Krige 2022).

However, this doesn't strike at the question of why BSxAI, and why *now.* Considerable policy attention has been expended on the potential misuse of biological technologies over the last 25 years (Evans 2025), but it is rare to see such a rapid and *public* development of interest by policymakers and practitioners—up to and including legislative proposals. A large part of the answer is undoubtedly the COVID-19 pandemic and contemporaneous societal changes, grounding our corpus as the next stage in the history of anxieties around biomedicine, infection and contagion provoking aggressive securitization (Wald 2008). The history of biosecurity is, in part, a history of new developments (be they societal or technological) leading to new framings of risk (see Lakoff 2017, particularly the discussion of the "generic biological threat"). So too is how heavily hyped both biotechnology and AI are as sources of imagined futures: risk and fear, as Nik Brown (2003) reminds us, are often predicated on (and an inevitable consequence of) hype. Taken together, it is unsurprising that

something that brings different domains of high technology into collision around viruses, diseases and contagion would prove a potent narrative.

Yet even the best framing in the world is of little use if a target audience does not encounter or actively work on the issues. This is where the *practice* comes into play. Far from being either an organic evolution of state and private interest, or a simple matter of convergent framing, the increased attention to BSxAI issues by policymakers is something that has been purposefully enculturated by the TESCREAL-aligned actors involved. The "Responsible AI x Biodesign" manifesto, for example—signed and authored by myriad actors in this space, and framed as a "community statement" —was in fact not only authored and advertised strategically, but explicitly written and timed to intervene in international policy discussions, with members of the (OPP-funded) research team that hosted and initiated the statement privately discussing having "quietly passed [the draft]" to figures in the United Kingdom and United States governments "to gauge...interest in lifting as a G7 deliverable" (Hebbeler 2024).

Alongside both research funding and direct pressure on state actors, TESCREAL-aligned figures have been explicit about encouraging movement adherents to pursue roles in government advising, with some, such as Jason Matheny (previously of IARPA, now the President of the RAND Corporation) directly lecturing on "roles for EAs working within government", pointing to movement-aligned organisations such as 80,000 Hours (founded by EA figure Will McAskill, and funded by OPP) as effective avenues to engaging with governmental work. This approach and encouragement seems to have paid off, with many of the civil service figures involved in biosecurity and AI originating in the EA movement, as documented in our Supplementary Data. At the AI Safety Institutes of the United Kingdom and United States, respectively – the examples with which we opened our paper – employees have moved there from (or to) the Centre for Long-Term Resilience, the Future of Humanity

Institute, CHS, the Existential Risks Alliance, and Metaculus: each OPP-funded, EA-aligned organisations in the private sector, epitomising Matheny's advice to EA adherents to engage in "continuous cycling throughout a career, bouncing around between the different sectors in order to bring knowledge across them" (Matheny 2017).

In effect, EA and aligned movements have developed an *independent field* that has a complete pipeline of advancement opportunities from early undergraduate through to senior positions. They fund their own think tanks and fellowships, or sponsor units within legacy think tanks such as RAND at the Nuclear Threat Initiative that align with their interests; recruit undergraduates through university societies, conferences, and directed engagement strategies; contribute to key graduate programs; build large online and physical communities; and generate their own formal and informal literatures. These products are then placed into the hands of ideologically aligned—and, increasingly, intentionally placed—actors within the private and public sector, from the boards of frontier artificial intelligence companies to government agencies.

This is not a unique strategy to EA; one could say the same thing about traditional think tanks, such as RAND Corp, the American Enterprise Institute, or the Heritage Foundation (Medvetz 2012). But it significant in the speed of its emergence. In 2019, an article in the *Bulletin of the Atomic Scientists* asked the question "Will splashy philanthropy cause the biosecurity field to focus on the wrong risks?" (Lentzos 2019), with a particular focus on the infusion of money into biosecurity by OPP. Five years later, a cluster of funders aligned with and centred around OPP exert a dominant position in biosecurity, to the point that they can mobilize resources to create novel issues, and insert them into public policy, whole cloth. This influence, moreover, extends to the funding of numerous think tanks, from legacy institutions such as RAND through to relative newcomers such as the Council on Strategic Risks. Concerns about the national security impact of computational approaches to biology

are two decades old. But BSxAI does not continue that discussion, instead emerging as a discrete literature and field that did not *exist* prior to its development by the field of actors created and funded by EA organisations. Its existence, moreover, has been characterized by an expensive, cradle-to-grave strategy that has generated a scholarly corpus that defines an issue; generated consensus statements on that issue; and then moved that consensus into the halls of power in the UK and US in less than five years—with the majority of that work arising in an 18-month window from 2023-2024. This, then, is the ultimate answer to how BSxAI reached its position of prominence in narratives around AI (and national) security: through the purposeful and ideologically-driven *creation* of a new domain of researcher speculation, with policymaking to match.

Further evidence for BSxAI being shaped by ideology and policy desires can be seen from how its framing has changed over time. While, as we note, this corpus begins heavily integrated in the language and literature of the TESCREAL bundle, with a particular focus on EA and longtermism, much of that language becomes denuded by the time the end of our data collection in 2024. In literature on BSxAI produced directly by CHS while supported with funding from OPP, no mention of GCBRs is found, despite their generation of that term only a few years before the corpus' emergence. We find it significant, given that it was CHS that developed the topic to begin with, that their preferred phrase is one found in O'Brien and Nelson: a "high-consequence" biological risk or event.

Why would GCBRs, despite their important role in beginning this corpus, be left out in its maturity? Here, we can only speculate. On the one hand, the primary author of the GCBR definitional paper no longer works within the field of TESCREAL-aligned and funded biosecurity, and this may simply reflect a natural thematic change in the field. We think this too quick, however, given the significant investment in GCR language by OPP and its fellow travellers over the latter period of the 2010s, which extend into the earliest papers in the

BSxAI corpus—but not further. Rather, we think the explanation may be one (or both) of the following framing strategies. First, "high-consequence" reads as non-ideological, while GCRs, existential risks, information hazards, and a range of other terms explicitly mark TESCREAL discourse. The transition into policymaking requires language that is intelligible to legislative bodies. "High-consequence" is a Washington or Westminster-compatible phrase, one that generates urgency but allows for its positioning among a range of other priorities.

Simultaneously, the avoidance of explicitly-EA tied language could be a response to effective altruism receiving considerable negative press in recent years. The arrest of Sam Bankman-Fried, a purported effective altruist who had provided early funding for OPP, led to a rash of publications about the politics of these movements, and their place in civil society. Lief Wenar (2024a), writing in *Wired*, used the event to illustrate and shed light on the presence of billionaires as prominent funders of EA organisations, and the sheer scope of EA power, from their recruitment of undergraduates to the funding of university departments and old-school think tanks. Taken together, he treated them as evidence of the capture of these institutions by the priorities of the ultra-wealthy who have contributed tens of billions of dollars to the cause (Todd 2021). A string of additional controversies, the most prominent of which was the shuttering of FHI after Nick Bostrom was revealed to have made several explicitly eugenic and anti-black racist comments, followed (Anthony 2024). It may be that abandoning explicit EA-charged language benefits this scholarship given that effective altruism was simultaneously becoming associated with, among other things, the worstexcesses of the technocratic class.

Discussion and conclusion

In this paper, we set out to understand the emergence, and underlying structure, of a literature concerning the biosecurity implications of AI. We outline a corpus of 73 documents produced between 2020 and 2024, that reveal a pattern of publication by individuals and groups funded by a common set of ideologically-aligned funders. Further investigation detailed how this group have created, from its inception, a field that aligns intellectual trends, financial resources, and personnel into a common movement that has elevated an issue from a series of hypotheticals into a core part of the ongoing debate over the governance of frontier AI.

For some readers, this may be interesting as a purely descriptive exercise – a mapping of a nascent, but increasingly-prominent, area of academic and public focus within biosecurity. But there is more to attend to here than simply the descriptive.

What we have documented is not simply the organic creation of a new field of scholarship, but the very purposeful problematisation of AI and biosecurity – based largely on the ideological priorities and priors of a narrow range of funders. These ideological priorities have shaped both the language and practices within BSxAI scholarship, constructing a novel domain of risk and concern. The corpus does so, moreover, within a very intentional framing towards policymaking in developed nations, that is attentive to the wider strategic hopes those nations have around AI, an industry that dominates its economic futures at the cost of other social programs.

This pipeline, while bearing some passing similarities to previous work in biosecurity, and while integrating with legacy institutions within the field, has enabled the generation of what is essentially a new literature cut from within the structure of the old. Our corpus, as near as we can tell, does little to interact with the previous literature in biosecurity. But it makes use of high-profile groups within the community and their pre-existing institutional power to

mobilize resources around this issue. Put in the parlance of biosecurity itself: TESCREAL aligned funders have borrowed the tacit knowledge—particularly the unwritten, practice-based, ostensive knowledge (Vogel 2006; Revill and Jefferson 2014)—of the biosecurity class to generate a new problem wholesale.

Further, we have seen the construction of a pipeline to convey these ideas of risk – and the attendant urgency of action – to policymakers, driven by the same funders ultimately responsible for the academic and think-tank scholarship evidencing the area. This pipeline has transformed BSxAI into a technical policymaking language, created ways for policymakers to access appropriate aligned scholars, and convened or pre-empted events of note to ensure the message of the field is in the right hands at the right time. It is not a singular organization pursuing this, but rather a connected web of organizations that are funded by a small number of ideologically aligned groups.

This should cause us to view the space with some scepticism. As Naomi Oreskes (2020) writes in reviewing the scandal of Jeffrey Epstein's funding of research at Harvard, the capacity for funders to hold sway over a field "undermines the integrity of the research enterprise when individuals can pick and choose lines of inquiry that appeal to them simply because they can pay for them." A narrow and ideologically-driven range of funders increases the risk of a field being subject to "epistemic capture", in which ways of knowing (and the things to be known) are shaped and constrained by a particular range of interests.

At a time of increasing funding scarcity, this case study is a warning to the biological sciences – and the fields that abut, interpolate and translate them – of the scientific cost that can come with resorting to funding from powerful groups that lack democratic oversight and are unresponsive to broader social priorities. Biosecurity was, in the accounts of its own scholars (Lentzos 2019), somewhat listless in the late 2010s with the exit of the Alfred P

Sloan Foundation and the Wellcome Trust from regular, broad funding opportunities for the field. There is a sense, then, in which a broad class of well-intentioned actors in a technical discipline with little opportunity for growth would view the arrival of a new, well-endowed funder as a huge windfall. We suspect that in 2026, with so many federal research grants in the USA in the wind, there are many fields that would jump at the chance to find a sympathetic philanthropic home.

The arrival of OPP in biosecurity—initially through the work of program officer at OPP who was a former US Department of Defense and Department of Health and Human Services employee, an ELBI fellow, and who greenlit (and later ran) the OPP-funded Nuclear Threat Initiative's biological risk portfolio—has created a situation in which a narrow set of priorities have dominated the space. In a field in which the most basic experimental evaluation of laboratory biosafety techniques goes unfunded, a single funder without public accountability for its priorities can direct researchers' attention as it chooses. As federal funding for the life sciences and public policy research continues to erode, organisations such as OPP present themselves a critical resource for scholars from graduate students through to established PIs. However, we worry that, in same the way the BSxAI is a "science that celebrates itself," the rise of venture philanthropy in the life sciences and public policy will create fewer, and less diverse intellectual contributions. The result will be more brittle, and blinkered, policymaking around biosafety, and (correspondingly and paradoxically) a higher likelihood of harms.

For scholars examining ideological aspects of science in our current late-capitalist moment, this case study also demonstrates a need for broader inquiry into the consequences of the "TESCREAL bundle" of ideologies. The canonical site for such studies and ideologies is seen as AI research, particularly AI ethics (Ferri and Gloerich 2023), but our paper

demonstrates that the entities promoting these ideologies have cast a far wider net. Scholars who are critically investigating the costs and consequences of these ideas must do the same.

Bibliography

- Adams, C. J., Crary, A., and Gruen, L. (eds.). (2023). The good it promises, the harm it does: Critical essays on effective altruism. Oxford: Oxford University Press.
- Annas, G. J. (2010). Worst case bioethics: death, disaster, and public health. Oxford, UK: Oxford University Press.
- Anthony, A. (2024). 'Eugenics on Steroids': The Toxic and Contested Legacy of Oxford's Future of Humanity Institute. *The Observer*, April 28. https://www.theguardian.com/technology/2024/apr/28/nick-bostrom-controversial-future-of-humanity-institute-closure-longtermism-affective-altruism, accessed 8 July 2026.
- Anthis, Jacy Reese. (2022) Some Early History of Effective Altruism, 8 July, https://jacyanthis.com/some-early-history-of-effective-altruism, accessed 15 May 2026.
- Attal-Juncqua, A., Cicero, A., Zhu, A. and Inglesby, T. (2026) AIxBio: opportunities to strengthen health security. Health Security, 24(1): 66-76.
- Baker, D. and Church, G. (2024) Protein Design Meets Biosecurity. *Science* 383: 349 DOI:10.1126/science.ado1671.
- Beckstead, N. (2015) The long-term significance of reducing global catastrophic risks, The GiveWell blog, August 13, http://blog.givewell.org/2015/08/13/the-long-term-significance-of-reducing-global-catastrophic-risks accessed 15 May 2026.

- Bloomfield, D., Pannu, J., Zhu, A.W., Ng, M.Y., Lewis, A., Bendavid, E., Asch, S.M., Hernandez-Boussard, T., Cicero, A. and Inglesby, T., 2024. AI and biosecurity: The need for governance. Science, 385(6711): 831-833.
- Boenig-Liptsin, M. and Hurlbut, J. (2016) Technologies of Transcendence at Singularity University. In: J. Hurlbut and H. Tirosh-Samuelson (eds.) Perfecting Human Futures. Wiesbaden: Springer, pp. 239-267.
- Bostrom, N. (2002). Existential risks: Analyzing human extinction scenarios and related hazards. Journal of Evolution and technology 9(1): 1-30.
- Bostrom, N. and Cirkovic, M.M. (eds.) (2008) Global catastrophic risks. Oxford, UK: Oxford University Press.
- Bostrom, N. (2011). Information hazards: A typology of potential harms from knowledge. Review of Contemporary Philosophy (10): 44-79.
- Bostrom, N. (2013) Existential risk prevention as global priority. Global Policy 4(1): 15-31.
- Brown, N., Rappert, B. and Webster, A. (2000) Introducing contested futures: From looking into the future to looking at the future. In: N. Brown, B. Rappert and A. Webster (eds.) Contested futures: A sociology of prospective techno-science. Farnham: Ashgate, pp. 3-20.
- Brown, N. (2003) Hope against hype-accountability in biopasts, presents and futures. Science & Technology Studies 16(2): 3-21.
- Buzan, B., Wæver, O., & De Wilde, J. (1998) Security: A new framework for analysis. Boulder, CO: Lynne Rienner Publishers.
- Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.M., Zietz, M., Hoffman, M.M. and Xie, W. (2018)

Opportunities and obstacles for deep learning in biology and medicine. Journal of the royal society interface, 15(141); doi:10.1098/rsif.2017.0387.

- Clarke, S., & Whittlestone, J. (2022) A survey of the potential long-term impacts of AI: how AI could lead to long-term changes in science, cooperation, power, epistemics and values. In: proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society; 19-21 May, Oxford, UK. New York, NY: Association for Computing Machinery, pp. 192-202.

- Daniels, M., & Krige, J. (2022) Knowledge regulation and national security in postwar America. Chicago: University of Chicago Press.

- DiEuliis, D., Ellington, A. D., Gronvall, G. K., & Imperiale, M. J. (2019) Does biotechnology pose new catastrophic risks?. In: T. Inglesby and A. Adalja (eds.) Global Catastrophic Biological Risks. Princeton, NJ: Springer, pp. 107-119.
- Dzau, V. (2026) Panic, Neglect, Repeat: The 2026 Ebola Outbreak Shows Us We Must Change the Pattern for Global Pandemic Preparedness and Response, National Academy of Medicine, May 25, 2026, https://nam.edu/perspectives/2026-ebola-outbreak-pandemic-preparedness/ accessed May 25, 2026.
- Edwards, B. (2019) Insecurity and emerging biotechnology: Governing misuse potential. Princeton, NJ: Springer.
- Enemark, C. (2017) Biosecurity dilemmas: dreaded diseases, ethical responses, and the health of nations. Washington, DC: Georgetown University Press.
- Evans, N. G., and Selgelid, M. J. (2015) Biosecurity and open-source biology: the promise and peril of distributed synthetic biological technologies. Science and engineering ethics 21(4): 1065-1083.

- Evans, N.G. (2016) Ebola: From Public Health Crisis to National Security Threat. In F. Lentzos (ed.) Biological Threats in the 21st Century. London, UK: Imperial College Press edited by Filippa Lentzos. Imperial College Press, pp. 277-292.
- Evans, N.G. (2025) Gain of Function. Cambridge, MA: MIT Press.
- Ferri, G., & Gloerich, I. (2023) Risk and harm: Unpacking ideologies in the AI discourse. In: Proceedings of the 5th International conference on Conversational User Interfaces; 19-21 July, Eindhoven, NL. New York, NY: Association for Computing Machinery, 28: 1-6.
- G7. (2023) G7 Hiroshima process on generative artificial intelligence (AI) towards a G7 common understanding on generative AI. G7, 7 September 2023, https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/09/g7-hiroshima-process-on-generative-artificial-intelligence-ai_8d19e746/bf3c0c60-en.pdf, accessed 8 July 2026.
- Garfinkel, M. S., Endy, D., Epstein, G. L., and Friedman, R. M. (2007) Synthetic genomics: options for governance. Industrial Biotechnology 3(4): 333-365.
- Gebru, T., & Torres, É. P. (2024). The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence, First Monday, 29(4); doi: 10.5210/fm.v29i4.13636.
- Gleiberman, M. (2023) Effective altruism and the strategic ambiguity of 'doing good'. Antwerp, BE: University of Antwerp Institute of Development Policy and Management.
- Götting, J., Medeiros, P., Sanders, J.G., Li, N., Phan, L., Elabd, K., Justen, L., Hendrycks, D. and Donoughe, S. (2025) Virology capabilities test (VCT): a multimodal virology Q&A benchmark. *Preprint at https://arxiv. org/abs/2504.16137*.

- Greaves, H. and McAskill, W. (2021) The case for strong longtermism. Oxford, UK: Global Priorities Institute. GPI Working Papers 5-2021.
- Guillemin, J. (2006) Biological Weapons: From the Invention of State-Sponsored Programs to Contemporary Bioterrorism. New York, NY: Columbia University Press.
- Hebbeler, A. (2024) Andrew. Email re: Community guidelines finished?? (AI+protein design). Seattle, WA: Institute for Protein Design, 8 February.
- Her Majesty's Government (HMG) (2023) UK Biological Security Strategy, 12 June, https://www.gov.uk/government/publications/uk-biological-security-strategy/uk-biological-security-strategy-html, accessed 9 July 2026.
- Jeffery, N., Carter, S.R., Alexanian, N., Crook, O., Curtis, S., Moulange, R., Rath, S., Rose, S. and Clark, J. (2023). Bio X AI: Policy recommendations for a new frontier. Federation of American Scientists blog, December 12, https://fas.org/publication/bio-x-ai-policy-recommendations/, accessed 9 July 2026.
- Johnson, A. (2009) Hitting the brakes: Engineering design and the production of knowledge. Durham, NC: Duke University Press.
- Keyes, O. (2020) Automating autism: Disability, discourse, and artificial intelligence. The Journal of Sociotechnical Critique 1(1): 8-39.
- Koblentz, G.D. (2017) The de novo synthesis of horsepox virus: implications for biosecurity and recommendations for preventing the reemergence of smallpox. Health security, 15(6): 620-628.
- Lakoff, A. (2017) Unprepared: global health in a time of emergency. Oakland, CA: University of California Press.
- Leitenberg, M. (1999) Aum Shinrikyo's efforts to produce biological weapons: A case study in the serial propagation of misinformation. Terrorism and Political Violence 11(4): 149-158.

- Leitenberg, M. and Zilinskas, R. (2012) The Soviet Biological Weapons Program. Cambridge, MA: Harvard University Press.
- Lentzos, F. (2009) The Pre-History of Biosecurity: Strategies of Managing Risks to Collective Health. In: B. Rappert and C. Gould (eds.) Biosecurity: Origins, Transformations and Practices. London: Palgrave Macmillan, pp. 25-41.
- Lentzos, F. (2019) Will splashy philanthropy cause the biosecurity field to focus on the wrong risks?, April 25, Bulletin of the Atomic Scientists, https://thebulletin.org/2019/04/will-splashy-philanthropy-cause-the-biosecurity-field-to-focus-on-the-wrong-risks/, accessed 9 July 2026.
- Loader, I. and Walker, N. (2007) Civilizing Security. Cambridge, UK: Cambridge University Press.
- Long, C. M., & Marzi, A. (2021) Biodefence research two decades on: worth the investment?. The Lancet Infectious Diseases 21(8): e222-e233.
- MacAskill, W. (2017) Effective altruism: introduction. Essays in Philosophy 18(1): 1-5.
- MacAskill, W. (2019) The Definition of Effective Altruism. In: H. Greaves and T. Pummer (eds.) Effective Altruism: Philosophical Issues. Oxford, UK: Oxford University Press.
- MacKenzie, D. (1993) Inventing accuracy: A historical sociology of nuclear missile guidance. Cambridge, MA: MIT Press.
- Marijan, B. (2023) The dilemma of dual-use AI. Project Ploughshares, 18 September, https://ploughshares.ca/the-dilemma-of-dual-use-ai/, accessed 9 July 2026.

- Matheny, J. (2017) Effective altruism in government. Effective Altruism forum, 2 June, https://forum.effectivealtruism.org/posts/arzY8HSfC4PtfxA4w/jason-matheny-effective-altruism-in-government, accessed 9 July 2026.
- Matthews, D. (2015) I Spent a Weekend at Google Talking with Nerds about Charity. I Came Away…Worried, Vox, 10 August, https://www.vox.com/2015/8/10/9124145/effective-altruism-global-ai, accessed 9 July 2026
- Matthews, D. (2018) You have $8 billion. You want to do as much good as possible. What do you do?, Vox, 16 October, https://www.vox.com/2015/4/24/8457895/givewell-open-philanthropy-charity, accessed 9 July 2026
- Medvetz, T. (2012) Think Tanks in America. Chicago, IL: Chicago University Press.
- Nash, D. (2024) 2024 July EA Updates, Monthly Overload of Effective Altruism, 1 July, https://moea.substack.com/p/2024-july-ea-updates, accessed 9 July 2026
- National Research Council (2006) Globalization, Biosecurity, and the Future of the Life Sciences. Washington, DC: The National Academies Press.
- National Security Comission for Emerging Biotechnologies (NSCEB) (2024) AIxBio White Paper 1: Introduction to AI and Biotechnology. Washington, DC: National Security Comission for Emerging Biotechnologies. NSCEB White Paper Series on AIxBio no. 1.
- National Security Comission for Emerging Biotechnologies (NSCEB) (2025) Charting the Future of Biotechnology: An action plan for American security and prosperity. Washington, DC: National Security Comission for Emerging Biotechnologies. NSCEB Final Report.

- Office of the Director of National Intelligence (ODNI) (2024) Annual Threat Assessment of the U.S. Intelligence Community. Washington, DC: Office of the Director of National Intelligence.
- Oreskes, N. (2020), Jeffrey Epstein's Harvard Connections Show How Money Can Distort Research, Scientific American, 1 September, https://www.scientificamerican.com/article/jeffrey-epsteins-harvard-connections-show-how-money-can-distort-research/, accessed 9 July 2026.
- Ouagrham-Gormley, S. B. (2014) Barriers to bioweapons: the challenges of expertise and organization for weapons development. Ithaca, NY: Cornell University Press.
- O'Brien, J. T., & Nelson, C. (2020) Assessing the risks posed by the convergence of artificial intelligence and biotechnology. Health security, 18(3): 219-227.
- Pannu, J., Bloomfield, D., Zhu, A., MacKnight, R., Gomes, G., Cicero, A. and Inglesby, T.V. (2024) Prioritizing high-consequence biological capabilities in evaluations of artificial intelligence models. arXiv preprint arXiv:2407.13059.
- Rabinow, P. & Bennett, G. (2012) Designing Human Practices: An Experiment with Synthetic Biology. Chicago, IL: University of Chicago Press.
- Rappert, B. (2007) Biotechnology, Security and the Search for Limits: An Inquiry into Research and Methods. London: Palgrave Macmillan.
- Responsible AIxBiodesign (2023) Community Values, Guiding Principles, and Commitments for the Responsible Development of AI for Protein Design, Institute for Protein Design, 8 March, https://responsiblebiodesign.ai/, accessed 9 July 2026
- Revill, J. and Jefferson, C. (2014). Tacit knowledge and the biological weapons regime. Science and Public Policy, 41(5): 597-610.

- Rose, N. (1998) Inventing our selves: Psychology, power, and personhood. Cambridge, UK: Cambridge University Press.
- Sandberg, A. and Nelson, C. (2020) Who should we fear more: biohackers, disgruntled postdocs, or bad governments? A simple risk chain model of biorisk. Health security 18(3): 155-163.
- Sandbrink, J. (2023) Panoptic dual-use management: preventing deliberate pandemics in an age of synthetic biology and artificial intelligence. Ph.D. Dissertation, Oxford University, Oxford, UK.
- Sandifer, E. (2017) Neoreaction a Basilisk: Essays on and around the Alt-Right. Rochester, NY: Eruditorum Press.
- Schoch-Spana, M., Cicero, A., Adalja, A., Gronvall, G., Kirk Sell, T., Meyer, D., Nuzzo, J.B., Ravi, S., Shearer, M.P., Toner, E. and Watson, C. (2017) Global catastrophic biological risks: toward a working definition. Health Security 15(4): 323-328.
- Selgelid, M.J. and Weir, L. (2010) Reflections on the synthetic production of poliovirus. Bulletin of the Atomic Scientists, 66(3): 1-9.
- Smith, J.A. and Sandbrink, J.B. (2022) Biosecurity in an age of open science. PLoS biology 20(4): e3001600.
- Smith, N. (2026), New security unit to tackle ‘catastrophic’ AI-enabled bioweapons, Australian Financial Review, 24 April, https://www.afr.com/politics/federal/new-security-unit-to-tackle-catastrophic-ai-enabled-bioweapons-20260424-p5zqvw, accessed 9 July 2026
- Soice, E.H., Rocha, R., Cordova, K., Specter, M. and Esvelt, K.M. (2023) Can large language models democratize access to dual-use biotechnology?. arXiv preprint arXiv:2306.03809.

- Syme, T. (2019) Charity vs. revolution: effective altruism and the systemic change objection. Ethical theory and moral practice 22(1): 93-120.
- Talinn, J. (2018) Fireside Chat, Effective Altruism, 8 July, https://www.effectivealtruism.org/articles/ea-global-2018-jaan-fireside-chat, accessed 9 July 2026
- Todd, B. (2021) Is Effective Altruism growing?, 80,000 Hours, 28 July, https://80000hours.org/2021/07/effective-altruism-growing/, accessed 9 July 2026
- Tucker, J.B. (2000) Toxic Terror: Assessing Terrorist Use of Chemical and Biological Weapons. Cambridge, MA: MIT Press.
- Tucker, J.B. (2010) Double-edged DNA: Preventing the misuse of gene synthesis. Issues in Science and Technology 26(3): 23-32.
- Tucker, J.B. (2012) Innovation, Dual Use, and Security: Managing the Risks of Emerging Biological and Chemical Technologies. Cambridge, MA: MIT Press.
- Turchin, A. and Denkenberger, D. (2020) Classification of global catastrophic risks connected with artificial intelligence. AI & Society 35(1): 147-163.
- US AI Safety Institute and UK AI Safety Institute (2024) US AISI and UK AISI Joint Pre-Deployment Test: Anthropic's Claude 3.5 Sonnet. Washington, DC: National Institutes of Standards and Technology.
- US Senate (2025) Strategy for Public Health Preparedness and Response to Artificial Intelligence Threats. United States Senate, 10 February, https://www.congress.gov/bill/119th-congress/senate-bill/501/all-actions, accessed 9 July 2026.
- Vogel, K. (2006) Bioweapons proliferation: Where science studies and public policy collide. Social Studies of Science 36(5): 659-690.

- Vogel, K. (2012) Phantom Menace or Looming Danger? A New Framework for Assessing Bioweapons Threats. Baltimore, MD: Johns Hopkins University Press.
- Waidzunas, T. and Epstein, S. (2015) 'For men arousal is orientation': Bodily truthing, technosexual scripts, and the materialization of sexualities through the phallometric test. Social Studies of Science 45(2): 187-213.
- Wald, P. (2008) Contagious: Cultures, carriers, and the outbreak narrative. Durham, NC: Duke University Press.
- Walsh, M. (2025) One of the world's most influential philanthropies is changing its name. Here's why it matters. Vox, 18 November, https://www.vox.com/future-perfect/469038/open-philanthropy-alexander-berger-coeffective-giving-effective-altruism, accessed 9 July 2026.
- Westerstrand, S., Westerstrand, R. and Koskinen, J. (2024) Talking existential risk into being: A Habermasian critical discourse perspective to AI hype. AI and Ethics 4(3): 713-726.
- Wenar, L. (2024a) The Deaths of Effective Altruism. Wired, 27 March, https://www.wired.com/story/deaths-of-effective-altruism/, accessed 26 September 2025
- Wenar, L. (2024b) To my young friends in EA, October, https://docs.google.com/document/d/1ytfQOfmjWTDiGdjuyNBfWJ_g-644y_Pd/edit, accessed 9 July 2026
- Wheelis, M., Rózsa, L. and Dando, M. (2000) Deadly Cultures: Biological Weapons Since 1945. Cambridge, MA: Cambridge University Press.
- Wilson, J. and Winston, A. (2024) Sam Bankman-Fried Funded a Group with Racist Ties. FTX Wants Its $5m Back. The Guardian, 16 June,

https://www.theguardian.com/technology/article/2024/jun/16/sam-bankman-fried-ftx-eugenics-scientific-racism, accessed 9 July 2026.

- Wright, S. (1986) Molecular biology or molecular politics? The production of scientific consensus on the hazards of recombinant DNA technology. Social Studies of Science 16(4): 593-620.
- Yudkowsky, E. (2008) Artificial Intelligence as a positive and negative factor in global risk. In: in Bostrom, N. and M.M. Cirkovic (eds.) Global Catastrophic Risks. Oxford, UK: Oxford University Press.
- Yudkowsky, E. (2016) The AI alignment problem: why it is hard, and where to start. Speech to Stanford Symbolic Systems Program, 5 May.

Tables and Figures

Funder	Appearances	TESCREAL-Aligned
Open Philanthropy Project/Coefficient Giving	12	Y
Effective Giving/Ergo Impact	3	Y
National Institutes of Health	3	N
Horizon Institute	2	Y
Long-Term Future Fund	2	Y
Fidelity Charitable	2	N
Sentinel Bio	1	Y
ML Alignment Theory Scholars Program	1	Y
Vitalik Buterin	1	Y
DALHAP (Do As Little Harm As Possible) Investments, Ltd	1	Y
Waking Up Foundation	1	Y
Jaan Tallinn	1	Y

Longview	1	Y
Good Ventures	1	Y
Fredrik Österberg	1	N
Founders Pledge	1	Y
Effektiv Spenden	1	Y
Chan-Zuckerberg Biohub	1	N
Stanford University	1	N
National Science Foundation	1	N
BlueDot Impact	1	Y

Table 1: Declared funders within the BSxAI corpus

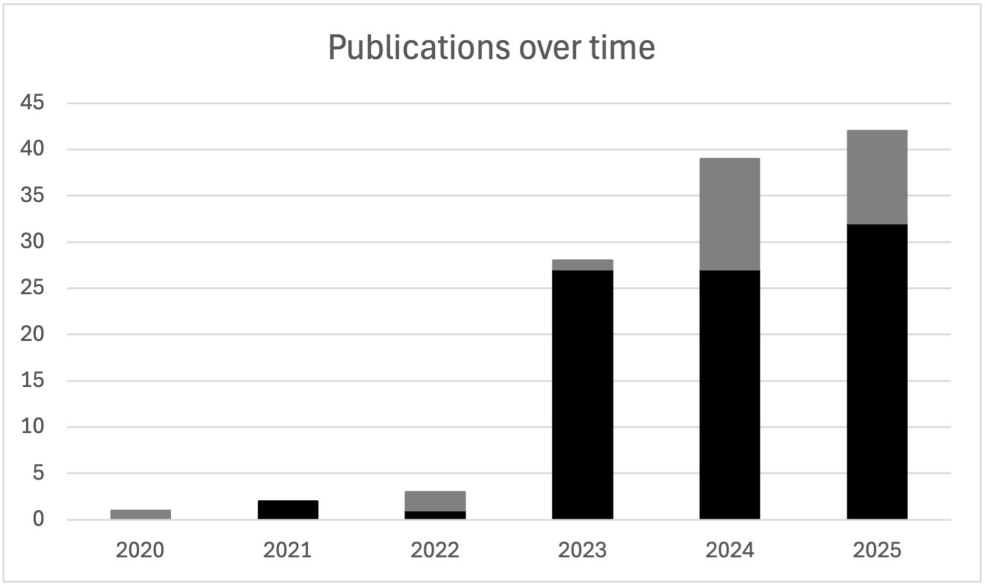


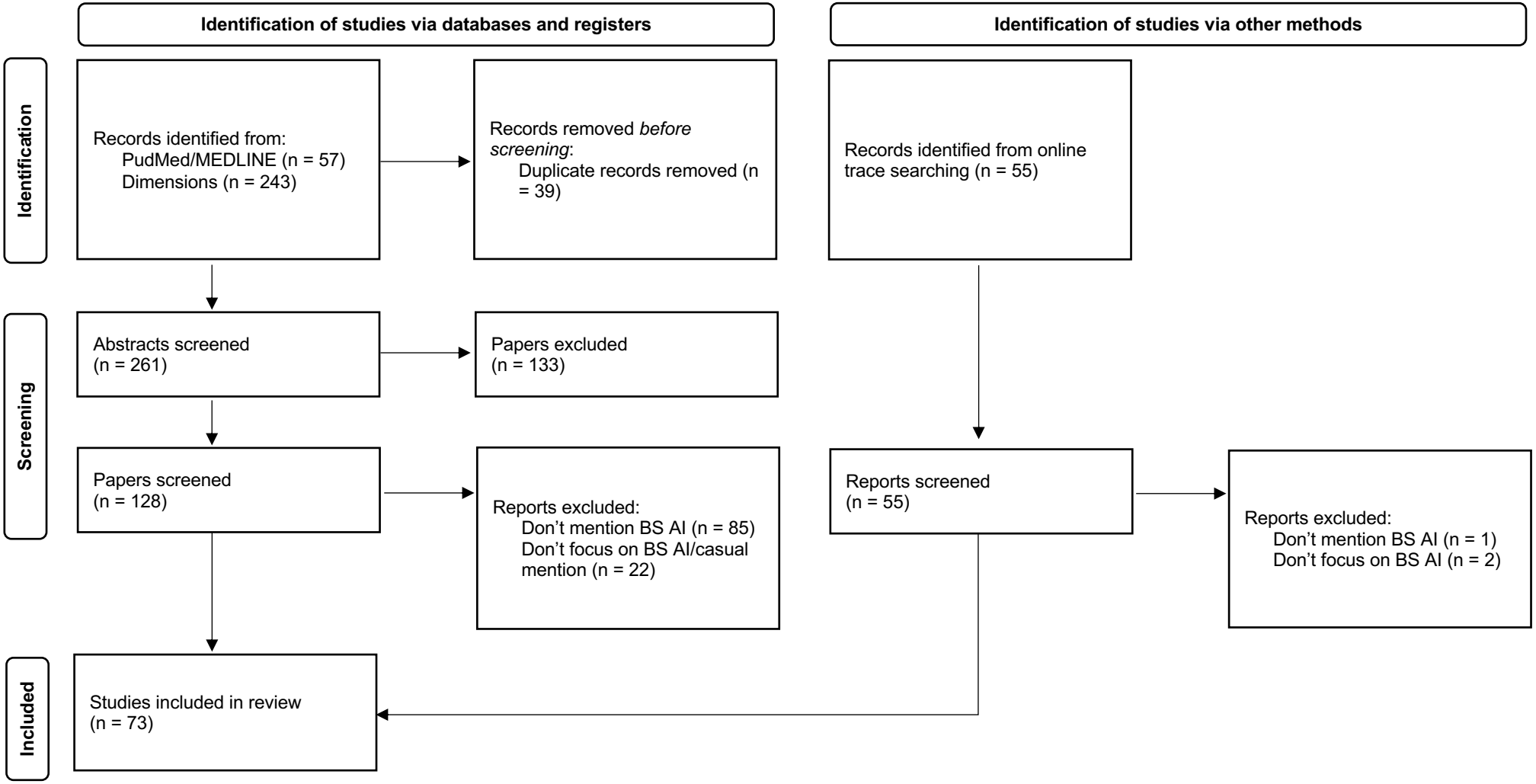
Figure 1 - publications over time. Grey = peer reviewed documents; black = non-peer reviewed documents

Appendices

Search strategies

NLM/PubMed - ((biosecurity[Title/Abstract] OR biorisk[Title/Abstract] OR biosafety[Title/Abstract] OR bioeconomy[Title/Abstract] OR biological weapons[Title/Abstract] OR biological risk[Title/Abstract]) AND (artificial intelligence[Title/Abstract] OR large language model[Title/Abstract] OR GPT[Title/Abstract]))

Dimensions.AI - “(("artificial intelligence") OR ("large language model") OR (GPT)) AND ((biosecurity) OR (biorisk) OR (biosafety) OR (bioeconomy) OR ("biological weapons") OR ("biological risk"))” Dimensions searches title/abstract as default, which is what was pursued here


Identification of studies via databases and registers
Identification of studies via other methods
Identification
Screening
Included
Records identified from:
PudMed/MEDLINE (n = 57)
Dimensions (n = 243)
Records removed *before screening*:
Duplicate records removed (n = 39)
Records identified from online trace searching (n = 55)
Abstracts screened
(n = 261)
Papers excluded
(n = 133)
Papers screened
(n = 128)
Reports excluded:
Don't mention BS AI (n = 85)
Don't focus on BS AI/casual mention (n = 22)
Reports screened
(n = 55)
Reports excluded:
Don't mention BS AI (n = 1)
Don't focus on BS AI (n = 2)
Studies included in review
(n = 73)

[i] Other origin stories exist for US and international biosecurity, including the Dalles salad bar attack in Oregon; South African Project COAST; use of chemical weapons in the Iraq-Iran War, latter careers of former Unit 731 members in Japan leading to contaminated blood supplies, and many more. For broader histories, see e.g. Tucker, 2000; Wheelis, Rózsa, and Dando (eds.), 2000; and Guillemin 2006.