On research oversight for fostering responsible AI R&D ecosystem

Neeti Pokhriyal

RAND Corporation
npokhriyal@rand.org

Abstract

An overlooked area in AI Governance is the discussion on oversight of research and development related to AI. We are inspired by the work on federal research oversight policies in bio-safety and bio-security communities, which are critical components of effective governance and ensure the responsible conduct of potentially dual-use research. In this paper, we highlight the need for discussions related to the critically important topic of research oversight of AI R&D and the potential benefits of such oversight to the AI R&D ecosystem. We identify if and how the frameworks in the bio-safety and bio-security community can help us think through the development of such a policy framework for AI, identify the critical challenges unique to AI when thinking of research oversight policies and procedures, and propose a suggested framework to mitigate some of the challenges. The intent of this paper is to ensure that there are appropriate risk identification and mitigating methods in place to prevent incidents and to raise awareness among researchers, academic institutions, and funding agencies about concerns related to AI safety and security.

Introduction

An overlooked area in AI Governance is the discussion on federal oversight of research and development related to AI. Rapid advances in AI technology and efforts to embed it in critical sectors like health, education, defense, etc., to provide societal benefits also pose new risks and have spurred efforts related to regulatory and nonregulatory mechanisms and guidelines for the governance of AI technologies. The focus of these efforts has been wide-ranging from aspects related to training for federal employees, disclosure of AI use by individual agencies, export controls, protection of privacy and worker's rights, risk management, and mitigation standards.

However, the focus of this paper is different. Many of the above efforts can be viewed as framing regulatory guidelines for *aposteriori* AI technologies, i.e., once the AI models are developed, deployed, and integrated across various sectors of society. Here, we focus on guidelines for *apriori* oversight of AI, i.e., questions that govern if the research has a dual use potential and, if yes, what are the appropriate risk mitigation steps in place that prevent undue harm. Dual-use research can be utilized for both benevolent and harmful purposes (National Institute of Health, 2023). Research oversight of AI R&D also encompasses questions related to increasing awareness and education of AI researchers, developers, their research institutions, and funding agencies.

As an exemplar, the biosecurity and biosafety communities have policies for research oversight, which are a critical component of effective governance and ensure the responsible conduct of dual-use research. The 2012 Federal DURC, the 2014 Institutional DURC, and the 2017 P3CO Framework policies are key components of the federal oversight framework for research in biosafety and biosecurity and have recently been modified in May 2024 as the United States Government (USG) Policy for Oversight of Dual Use Research of Concern and Pathogens with Enhanced Pandemic Potential. The policies have provided useful guidelines and awareness to researchers, academic institutions, and federal funding agencies on the critical safety and security concerns related to this type of research.

It is important to highlight that the types of risks that research oversight in biosafety and security communities deal with typically encompass research that could lead to the development of toxins or pathogens, increase their transmissibility to humans or society, or raise potential national security concerns, e.g., creation of chemical-biological weapons. It also encompasses security incidents, likely caused by misuse of knowledge or technology. Drawing an analogy, we are looking for similar types of risks in AI that oversight mechanisms might be able to prevent.

In this paper, we bring the idea of research oversight to the federal AI R&D ecosystem. We argue that if there are oversight guidelines and policies, which could be binding or nonbinding, that ensure that appropriate measures are in place within the academic or research institutions, researchers, and funding agencies in a manner that is commensurate with the risk to minimize adverse impacts on research innovation and openness to share the benefits of research.

Multiple stakeholders are affected by the policies and issues related to research oversight. These are 1) Federal funding agencies, which could be a federal department, agency, or office that funds research within the United States or internationally; 2) the principal investigators (PI); and 3) the academic or research institutions, which provide further oversight and ensure that the research is performed safely. The oversight framework assigns responsibilities to each of the stakeholders so that if a potential dual-use scenario is identified by the PI, their institution and funding agency can be informed/alerted, and necessary oversight mechanisms, such as setting up a committee for institutional review, coming up with risk mitigation plan, etc., can be invoked.

Our high-level contributions are as follows:

1. Highlight the need for research oversight of AI R&D,
2. Highlight the potential benefits of such an oversight to the AI R&D ecosystem,
3. Identify if and how the frameworks in the bio-safety and bio-security (henceforth referred to as life-sciences) community can help us think through the development of such a policy framework,
4. Identify the critical challenges unique to AI when thinking of research oversight policies and procedures
5. Propose a suggested framework to mitigate some of the challenges.

Challenges

A critical challenge is the definition of what constitutes dual-use AI research. Dual-use research can be utilized for both benevolent and harmful purposes (National Institute of Health, 2023). Most discussions on dual-use have been in the biosafety [1] and biosecurity [2] disciplines (National Academies of Sciences, Medicine et al. 2017; Policy 2024).

There's a smaller subset of research termed "Dual use research of concern (DURC)" in life sciences, which is defined as "research that, based on current understanding, can be reasonably anticipated to provide knowledge, information, products, or technologies that could be directly misapplied to pose a significant threat with broad potential consequences to public health and safety, crops and other plants, animals, the environment, materiel, or national security" (National Institute of Health, 2023). For research that is deemed DURC, governments around the world have a set of policies aimed at maximizing its benefits while minimizing the risk of misuse from knowledge or technologies provided by such research (Lev 2019; Himmel et al. 2019; Williams-Jones, Olivier, and Smith 2014).

We posit that rather than focusing on the dual-use of AI, we focus on a smaller subset of research in AI and term it as "dual-use research of concern" in AI, drawing an analogy from the life-sciences research that has grappled with similar issues. We argue that the concept of dual-use research of concern (DURC) in Artificial Intelligence (AI) has received very scant attention from the policymakers as well as the researcher community.

However, providing an equivalent definition of DURC in AI is a challenging task owing to the complexity involved in qualifying different factors, including, but not limited to, understanding if dual-use arises because of the model or the outputs of research. Very recent works have wrestled with finding a definition of dual-use for Natural Language Processing (Koplin 2023; Kaffee et al. 2023), or concerning Large Language Models (Grinbaum and Adomaitis 2023) or foundational models (Henderson et al. 2023), and do provide a good starting point for these discussions, but fall short of providing a comprehensive definition.

More concretely, our paper makes the following **contributions**:

1. We have provided a definition of DURC in AI and a risk-assessment framework as starting steps in this direction for building research oversight policies and procedures for AI.
2. We compare and contrast the policies and regulatory mechanisms for safeguarding DURC in the life sciences. We elaborate on the challenges that are common for the two disciplines, as well as highlight the ones that are unique to AI. We highlight the unique factors that are contributing to the potential for DURC.
3. We focus on how AI safety education, pedagogical tools, and training of researchers, especially early-stage students, can be instrumental in identifying, assessing, and mitigating the DURC criteria in AI.

The rest of the paper is organized as follows: we start by attempting a workable definition for DURC in AI; then we compare the learnings from DURC criteria in life sciences and contrast it with the specific challenges in AI; and, later, we present a risk assessment framework for researchers that can be employed as a starting point to initiate discussions on DURC and AI safety.

Defining Dual-use research of concern in AI

The term "dual-use" has historically been used in the context of defense applications and refers to technologies that have the potential to be used for peaceful and military purposes. While governments across the world have acknowledged the role of AI as a dual-use technology (Ueno 2023; Carrozza, Marsh, and Reichberg 2022).

As noted earlier, a precise definition of DURC in the context of AI is a key first step. The Biden administration's 2023 Executive Order (EO) (White House 2023) mentions the term "dual-use" for foundation models, as – "Dual-use foundation model means an AI model that is trained on broad data; generally uses self-supervision; contains at least tens of billions of parameters; is applicable across a wide range of contexts; and that exhibits, or could be easily modified to exhibit, high levels of performance at tasks that pose a serious risk to security, national economic security, national public health or safety, or any combination of those matters."

While it is up for debate, we believe that the above definition is both wide-ranging in terms of context but simultaneously restrictive, as it defines dual-use only in terms of foundation models. Why constrain the definition of "dual-use" only regarding foundation models? What about ML

[1] Biosafety deals with the application of practices, controls, and containment infrastructure that reduces the risk of unintentional exposure to, contamination with, release of, or harm from pathogens, toxins, and other associated biological materials.

[2] Biosecurity is the application of security measures designed to prevent the loss, theft, misuse, diversion, unauthorized possession or material introduction, or intentional release of pathogens, toxins, biological materials, and related information and/or technology.

models that do not qualify as foundation models but have been shown to raise dual-use concerns in designing toxins and new materials, as evidenced by (Urbina et al. 2022b; Shankar and Zare 2022). For instance, in a recent study (Urbina et al. 2022a), the authors showed how a traditional ML-based model for designing molecules intended for potential development to treat Alzheimer's disease could be easily modified to design toxins. Similarly, dual-use concerns for generative AI models for text and image generation (e.g., deepfakes (Jones 2023)) have received considerable attention as well (Koplin 2023; Grinbaum and Adomaitis 2023).

At the same time, a very broad definition (such as labeling all AI as potentially dual-use) could negatively impact the innovation in this field and will fail to realize the potential of AI as a catalyst for multiple and diverse domains, from scientific discovery (Wang et al. 2023) to psychology sciences (Demszky et al. 2023) to diplomatic decision making (Pokhriyal and Koebe 2023).

This is not a straightforward task, given that there is considerable ambiguity in defining dual-use even in life-sciences (Resnik 2009; Miller and Selgelid 2007; National Academies of Sciences, Medicine et al. 2017).

To the best of our knowledge, no current work in AI governance talks about a narrower but more risky subset of research, termed DURC in AI, and what such a classification would mean for the researchers, research institutions, funding agencies, and policy design. Here, we attempt to provide a definition of DURC for AI as follows: Dual-use research of concern for AI is any research that results in AI (trained model, technology, algorithm, products) that can be "easily adapted" by a malicious actor for harmful use.

As mentioned earlier, this definition intentionally goes beyond foundation models. One of the challenges in prescribing a workable definition of DURC in AI is the ambiguity in qualifying what AI itself means - is it the algorithms, methodologies, trained models, or technological artifacts that are driven by the core AI research?

Through the above definition, we stress the aspect of "potential for malicious use", so technical safeguards stating that a model's intended purpose and "Terms and Conditions" determining that the model need not be used for nefarious means and/or ethics statements do not work in this context, as used by industry standards.

It is critical to highlight the distinction between concerns of risks due to potential dual-use and other risks associated with AI, which includes ethical concerns (Bostrom and Yudkowsky 2014), issues related to bias and fairness (Zou and Schiebinger 2018; Mhasawade, Zhao, and Chunara 2021), and societal risks such as job loss due to automation (Shen and Zhang 2024), many of which are actively discussed in international venues such as the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society[3] and ACM Conference on Fairness, Accountability, and Transparency[4]. A related topic is the potential for malicious use of AI applications (Brundage et al. 2018) by exploiting security vulnerabilities of the application. These topics have been discussed but are separate from DURC and hence are not covered by the above definition.

Finally, it is debatable what it means for AI research to be "easily adapted" for malicious use. Foundation models make a strong case for such adaptability since they are designed to be easily fine-tuned for a target application. But can other AI models be adapted in the same way? As noted earlier, researchers have shown how a simple thought experiment could turn a beneficial drug-developing AI technology into a model that could generate potentially lethal chemicals.

Moreover, with the free availability of trained models and underlying data on public websites such as *HuggingFace*[5], the growing practice in the AI community to post their research on pre-print servers such as *arXiv*[6], and the increasing availability of low-cost computing and data infrastructure needed to develop AI models, it is clear that the threshold for "easy adaptability" is going to rapidly evolve, not to mention the hacks and workarounds to the guardrails that are put into place.

Unique challenges in AI, which need to be accounted for when considering research oversight.

In this section, we first briefly describe what DURC in life sciences means and the policies that have been put in place; then describe how some of the challenges of the AI community outlined in the previous section are similar to the challenges faced by life sciences community; and lastly detail some of the distinct challenges envisioned in conceptualizing DURC in AI.

DURC criteria in life sciences encompass research that involves a particular set of biological agents or toxins and certain types of experiments and is used in the context of biosecurity and biosafety (National Academies of Sciences, Medicine et al. 2017). For research deemed DURC, governments around the world have a set of policies aimed at maximizing its benefits while minimizing the risk of misuse from knowledge or technologies provided by such research (Lev 2019; Himmel et al. 2019; Williams-Jones, Olivier, and Smith 2014). These policies mainly cover institutional review and oversight of life sciences research, the roles and responsibilities of principal investigators or researchers, the role of funding agencies, and research institutions. The policies also provide requirements and performance standards for reviewing and publishing such research (National Academies of Sciences, Medicine et al. 2017).

As an example, in the U.S federal, policies are established by the National Institute of Health (NIH) and Health and Human Services to identify if there is a potential for DURC and to develop and implement risk mitigation measures for DURC, where applicable (National Institute of Health, 2023; Policy 2024; NIH Sourcebook; US Life Science DURC Policy), but similar policies do not yet exist for

[3] https://www.aies-conference.com/2023/

[4] https://facctconference.org/

[5] https://huggingface.co/

[6] https://arxiv.org/

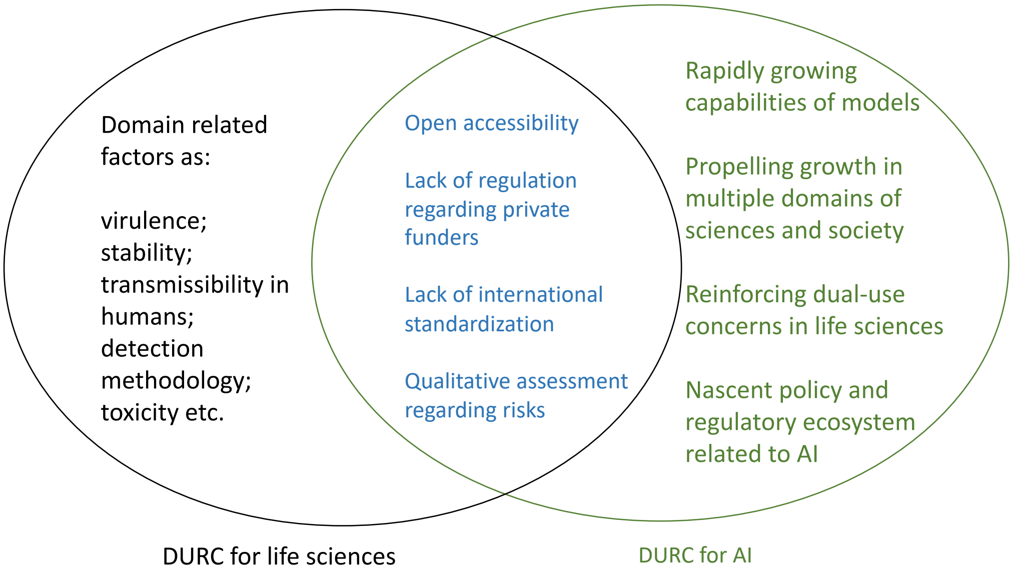
Domain related factors as:
virulence;
stability;
transmissibility in humans;
detection methodology;
toxicity etc.
Open accessibility
Lack of regulation regarding private funders
Lack of international standardization
Qualitative assessment regarding risks
Rapidly growing capabilities of models
Propelling growth in multiple domains of sciences and society
Reinforcing dual-use concerns in life sciences
Nascent policy and regulatory ecosystem related to AI
DURC for life sciences
DURC for AI

Figure 1: Contrasting implementing and interpreting DURC in life-sciences and AI. While there are some shared challenges faced by both communities, implementing DURC in AI has additional challenges and complexities, which need to be accounted for when thinking about research oversight.

AI research, at least in the U.S. We searched the policy documents for the U.S. National Science Foundation's Proposal & Award Policies & Procedures Guide (PAPPG) and NIH, two major funding agencies funding research in basic sciences and did not find policies regarding DURC in AI that can guide the potential principal investigators if such a concern arises.

For DURC criteria in life sciences, some risk mitigation frameworks are provided (Vennis et al. 2021; Tensmeyer et al. 2023); however, directly translating these frameworks in the AI domain is neither advisable nor straightforward. While there are some prominent risk management frameworks in AI - U.S. National Institute of Standards and Technology (NIST AI RMF), U.S. Department of Energy (DOE AI), Organisation for Economic Co-operation and Development (OECD) (OECD 2023) and the EU AI Act (EuropeanParliament 2023), none of them call out DURC in AI. The framework by OECD does mention dual use, but it is in the context of misuse and exploiting vulnerabilities.

We have attempted to illustrate the intersectionality of the DURC criteria both in life sciences and AI, as shown in Figure 1. Some common challenges are the tensions between making research openly accessible and safeguarding it so it cannot be misused, lack of regulation related to privately funded research, and qualitative risk assessments. The evolving nature of the threat landscape in life sciences concerns governments, who are pushing for tighter rules to safeguard research and establish appropriate guardrails and oversight (Heller 2023).

In life sciences, if access to knowledge or artifacts is controlled, the risk of DURC is somewhat mitigated. But what about open access with data and models that can be downloaded, and then fine-tuned, how do we control access then? Other challenges include wide-ranging applications of AI technologies (from societal tasks to scientific discovery) and their rapidly growing capabilities, coupled with nascent policies and regulatory ecosystems for safely conducting and deploying research.

Suggested framework to mitigate some of the challenges in thinking about research oversight in AI R&D

It is essential to establish mechanisms that guide the AI research community in identifying and assessing DURC-related risk and developing mitigating strategies. Central to these efforts is the need for an assessment framework that can be used by researchers to assess risks related to DURC in AI. A suggested framework is provided in Figure 2, where we define four primary dimensions for DURC identification and provide further sub-dimensions within each primary dimension.

The primary dimensions deal with assessing the potential of AI applications to be adapted for harmful use, who and what can be harmed, how easily it can be done, and whether it can result in newer risks and threats. We further divide each dimension into sub-dimensions to create a nuanced understanding and assessment of risks. For example, applications of DURC could span weapon design, breaking down social structures and jeopardizing the safety and reliability of critical infrastructures. The second dimension focuses on the extent of harm, focusing on vulnerable sections of our society, like children. The third dimension deals with ease of accessibility (via open-source, APIs, weights, and wider dissemination). The fourth dimension focuses on creating newer vulnerabilities in crucial areas, like precision medicine, and designing technologies that can evade detection. More dimensions can easily be incorporated.

The roles and responsibilities of different stakeholders

Developing educational modules in AI safety and security In recent years, the computer science education community has taken active steps to embed the concepts of ethics and responsible computing to educate students about the potential ethical implications of their work (Horton et al. 2022; Wong-Villacres et al. 2024), including efforts to incorporate ethics in AI education (Walsh et al. 2023). Similar developments will be needed to educate students, especially at early stages, to identify and assess AI safety risks. We again identify parallels with the life sciences community where the concepts related to DURC have been integrated into the curriculum (Miller and Selgelid 2007; Nixdorff 2013). We envision that such educational components would not only allow researchers to identify risks but also train them to develop mitigating strategies.

Role of policymakers An important role is for the policymakers to establish guidelines around DURC in AI, similar to frameworks set for life-sciences (National Academies of Sciences, Medicine et al. 2017; Revill, Husbands, and Bowman 2018). Such guidelines should ensure that the - Funding agencies - PI of AI researchers identify the potential risks with their research and adopt appropriate mitigation strategies. Recently developed frameworks, such as the NIST AI Risk Management Framework (AI RMF) (NIST AI RMF) in the US, have been set up to incorporate trust and safety considerations into AI development, and these frameworks could be extended to incorporate DURC using a framework similar to that suggested in Figure 2.

Role of academic institutions Implementation of such guidelines will require institutional units whose role will be similar to that of Institutional Review Boards (IRB) that can ensure that new research follows the established guidelines. Funding agencies that fund such research will need to update their guidelines to ensure that researchers identify any potential DURC-related risks in their proposed research.

Role of academic conferences and journals AI conferences and journals will need to update their review processes to ensure that researchers identify any potential DURC risks before submitting their manuscripts for consideration, similar to the Nature family of journals. AI/ML conferences and journals, require authors to provide a statement of the potential broader impact of their work, including potential ethical implications and future societal consequences and can be expanded to include DURC-related risks.

Benefits

Research oversight can also support the development of new ways to identify, manage, and mitigate serious risks due to AI systems, as well as research into AI failure cases, AI assurances, and AI incident databases. The researchers Developing AI safety education and pedagogical tools for researchers, especially in the early stages, in response to the DURC, will be in the best interest of the AI research community and the scientific enterprise.

We argue that one of the potential **benefits** of focusing on a narrower subset of research in AI, termed as DURC, is that it allows to focus on significant risks, indicating that the likelihood and magnitude of risks are such that they require careful assessment. Having a definition of DURC in AI also allows us to draw parallels in DURC as used in the life-sciences community - to learn from them and possibly adapt the robust policy frameworks already in place to identify and mitigate the risks arising from such concerns. Research that falls within the scope of this policy can increase our understanding AI-safety and security concerns.

Conclusions and Next steps

In this paper, we highlight the need for research oversight in AI R&D and borrow some ideas from such oversight policies already in the life-sciences community. We highlight the challenges and propose a framework that could be used as a guide in assessing risks.

Here, we define DURC in AI and discuss what can be learned from the robust mechanisms placed within the life-sciences community regarding this, as they have been grappling with dual-use issues for the past few decades. Once the research is identified as a potential for DURC in life sciences, it means that there are significant risks in conducting that research as it could be easily adapted for misuse. This translates to increased oversight for researchers, academic institutions, and funding agencies through oversights, review boards, and policy guidelines for safeguarding research so that its benefits can be maximized while minimizing the risks.

We highlight salient factors (distinct from the life-sciences community) that make assessment risk a particularly challenging task in AI. Last, we posit that researchers have a central role in building the criteria for DURC via educational programs and training focusing on AI safety in general and dual-use in particular. We also provide a risk assessment framework for guiding researchers in thinking about risks related to DURC in AI.

Our paper here is a first step in opening this conversation about the need for research oversight in AI-funded research. We have looked at the U.S. Federal funding ecosystem, but it would be interesting to see if such mechanisms exist in other countries. Also, we understand the deficiencies of DURC frameworks as the classification of ”research of concern” will always be subjective. More questions remain - what about research that is not federally funded; one of the aspects of DURC is that research is not publicly released - how to balance these concerns with open science ecosystem.

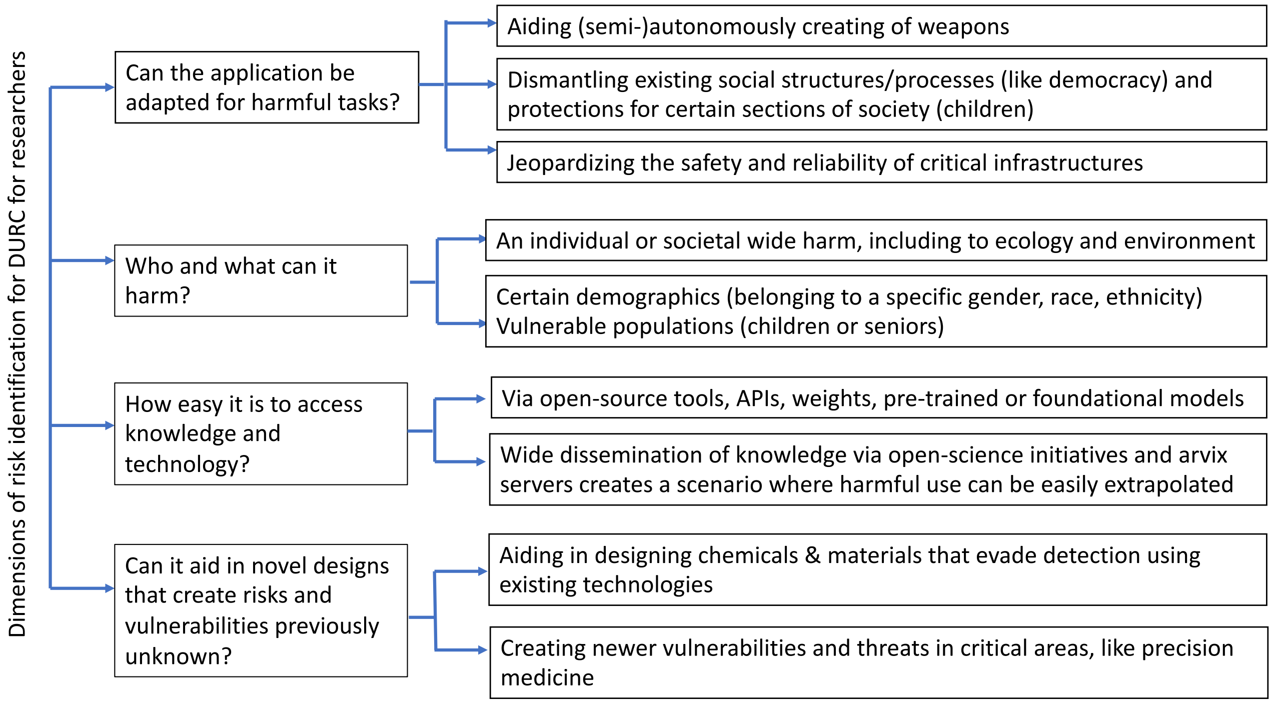
Dimensions of risk identification for DURC for researchers
Can the application be adapted for harmful tasks?
Aiding (semi-)autonomously creating of weapons
Dismantling existing social structures/processes (like democracy) and protections for certain sections of society (children)
Jeopardizing the safety and reliability of critical infrastructures
Who and what can it harm?
An individual or societal wide harm, including to ecology and environment
Certain demographics (belonging to a specific gender, race, ethnicity) Vulnerable populations (children or seniors)
How easy it is to access knowledge and technology?
Via open-source tools, APIs, weights, pre-trained or foundational models
Wide dissemination of knowledge via open-science initiatives and arvix servers creates a scenario where harmful use can be easily extrapolated
Can it aid in novel designs that create risks and vulnerabilities previously unknown?
Aiding in designing chemicals & materials that evade detection using existing technologies
Creating newer vulnerabilities and threats in critical areas, like precision medicine

Figure 2: Suggested framework to guide researchers in assessing risks related to DURC in AI. The 4 primary dimensions of the framework are given on the left, and each dimension is further subdivided into additional dimensions.

Acknowledgements

There was no funding available for this work.

References

Bostrom, N.; and Yudkowsky, E. 2014. *The ethics of artificial intelligence*, 316–334. Cambridge University Press.

Brundage, M.; Avin, S.; Clark, J.; Toner, H.; Eckersley, P.; Garfinkel, B.; Dafoe, A.; Scharre, P.; Zeitzoff, T.; Filar, B.; et al. 2018. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*.

Carrozza, I.; Marsh, N.; and Reichberg, G. M. 2022. Dual-Use AI Technology in China, the US and the EU. Technical report, Peace Research Institute Oslo (PRIO).

Demszky, D.; Yang, D.; Yeager, D. S.; Bryan, C. J.; Clapper, M.; Chandhok, S.; Eichstaedt, J. C.; Hecht, C.; Jamieson, J.; Johnson, M.; et al. 2023. Using large language models in psychology. *Nature Reviews Psychology*, 2(11): 688–701.

DOE AI. 2023. DOE AI Risk Management Playbook (AIRMP). Department, and of, and Energy.

European-Parliament. 2023. EU AI Act.

Grinbaum, A.; and Adomaitis, L. 2023. Dual Use Concerns of Generative AI and Large Language Models.

Heller, J. 2023. White House seeks input on tightening rules for risky pathogen research. *Science*.

Henderson, P.; Mitchell, E.; Manning, C.; Jurafsky, D.; and Finn, C. 2023. Self-Destructing Models: Increasing the Costs of Harmful Dual Uses of Foundation Models. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, 287–296. Association for Computing Machinery. ISBN 9798400702310.

Himmel, M.; et al. 2019. Emerging dual-use technologies in the life sciences: Challenges and policy recommendations on export control.

Horton, D.; McIlraith, S. A.; Wang, N.; Majedi, M.; McClure, E.; and Wald, B. 2022. Embedding Ethics in Computer Science Courses: Does it Work? In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1*, 481–487.

Jones, N. 2023. How to stop AI deepfakes from sinking society - and science. 621(7980).

Kaffee, L.-A.; Arora, A.; Talat, Z.; and Augenstein, I. 2023. Thorny Roses: Investigating the Dual Use Dilemma in Natural Language Processing. *arXiv preprint arXiv:2304.08315*.

Koplin, J. J. 2023. Dual-use implications of AI text generation. *Ethics and Information Technology*, 25(2): 32.

Lev, O. 2019. Regulating dual-use research: Lessons from Israel and the United States. *Journal of Biosafety and Biosecurity*, 1(2): 80–85.

Mhasawade, V.; Zhao, Y.; and Chunara, R. 2021. Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence*, 3(8): 659–666.

Miller, S.; and Selgelid, M. J. 2007. Ethical and philosophical consideration of the dual-use dilemma in the biological sciences. *Science and engineering ethics*, 13: 523–580.

National Academies of Sciences, E.; Medicine; et al. 2017. *Dual use research of concern in the life sciences: current issues and controversies*. National Academies Press.

National Institute of Health, 2023. 2023. Dual-Use Research. https://oir.nih.gov/sourcebook/ethical-conduct/special-research-considerations/dual-use-research.

NIH Sourcebook. Accessed on 01/31/24. Dual-Use Research. https://oir.nih.gov/sourcebook/ethical-conduct/special-research-considerations/dual-use-research, NIH Office of Intramural Research.

NIST AI RMF. 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0).

Nixdorff, K. 2013. Education for Life Scientists on the Dual-Use Implications of Their Research: Commentary on "Implementing Biosecurity Education: Approaches, Resources and Programmes". *Science and engineering ethics*, 19: 1487–1490.

OECD. 2023. Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI.

Pokhriyal, N.; and Koebe, T. 2023. AI-assisted diplomatic decision-making during crises—Challenges and opportunities. *Frontiers in big Data*, 6: 1183313.

Policy, U. S. G. 2024. *United States Government Policy for Oversight of Dual Use Research of Concern and Pathogens with Enhanced Pandemic Potential*. United States Government.

Resnik, D. B. 2009. What is "dual use" research? A response to Miller and Selgelid. *Science and engineering ethics*, 15(1): 3–5.

Revill, J.; Husbands, J.; and Bowman, K. 2018. *Governance of Dual Use Research in the Life Sciences: Advancing Global Consensus on Research Oversight: Proceedings of a Workshop*. The National Academies Press.

Shankar, S.; and Zare, R. N. 2022. The perils of machine learning in designing new chemicals and materials. *Nature Machine Intelligence*, 4(4): 314–315.

Shen, Y.; and Zhang, X. 2024. The impact of artificial intelligence on employment: the role of virtual agglomeration. *Humanities and Social Sciences Communications*, 11(1): 122.

Tensmeyer, N.; Crawford, E.; Greene, D.; and Kazmierczak, M. J. 2023. The Role of Technology Risk Assessment Frameworks in Research. *Available at SSRN 4554618*.

Ueno, H. 2023. *Artificial Intelligence as Dual-Use Technology*, 7–32. Springer International Publishing.

Urbina, F.; Lentzos, F.; Invernizzi, C.; and Ekins, S. 2022a. Dual use of artificial-intelligence-powered drug discovery. *Nature Machine Intelligence*, 4(3): 189–191.

Urbina, F.; Lentzos, F.; Invernizzi, C.; and Ekins, S. 2022b. A teachable moment for dual-use. *Nature machine intelligence*, 4(7): 607–607.

US Life Science DURC Policy. Accessed on 01/31/24. United States Government Policy for Oversight of Life Sciences Dual Use Research of Concern. Biosafety-and-Biosecurity-Policy.

Vennis, I. M.; Schaap, M. M.; Hogervorst, P. A.; de Bruin, A.; Schulpen, S.; Boot, M. A.; van Passel, M. W.; Rutjes, S. A.; and Bleijs, D. A. 2021. Dual-use Quickscan: A web-based tool to assess the dual-use potential of life science research. *Frontiers in bioengineering and biotechnology*, 9: 797076.

Walsh, B.; Dalton, B.; Forsyth, S.; and Yeh, T. 2023. Literacy and STEM Teachers Adapt AI Ethics Curriculum. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(13): 16048–16055.

Wang, H.; Fu, T.; Du, Y.; Gao, W.; Huang, K.; Liu, Z.; Chandak, P.; Liu, S.; Van Katwyk, P.; Deac, A.; et al. 2023. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972): 47–60.

White House. 2023. Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government. https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

Williams-Jones, B.; Olivier, C.; and Smith, E. 2014. Governing 'dual-use' research in Canada: A policy review. *Science and Public Policy*, 41(1): 76–93.

Wong-Villacres, M.; Kutay, C.; Lazem, S.; Ahmed, N.; Abad, C.; Collazos, C.; Elbassuoni, S.; Islam, F.; Singh, D.; Mayeesha, T. T.; Ujakpa, M. M.; Zaman, T.; and Bidwell, N. J. 2024. Making Ethics at Home in Global CS Education: Provoking Stories from the Souths. *ACM J. Comput. Sustain. Soc.*, 2(1).

Zou, J.; and Schiebinger, L. 2018. AI can be sexist and racist - it's time to make it fair. *Nature*, 559.