Towards Closing the Gap Between Moral Justifications for Research and Actual Impacts

Dr. Ryan Blake Jackson, Colorado College

Blake is an assistant professor of computer science at Colorado College. His research is in human-robot interaction, and he is specifically interested in verbal noncompliance and clarification interactions, robot ethics, natural language generation, robot gender presentation, and mental models of robot identity.

Dr. Cory B Scott, Colorado College

I am an assistant professor in the Department of Mathematics and Computer Science at Colorado College. My research focuses on developing machine learning methods for computational biology and computational geometry.

Towards Closing the Gap Between Moral Justifications for Research and Actual Impacts

Abstract

The purpose of this WIP research paper is to discuss a pattern in computer science publications that we term "sidepromising," wherein researchers develop harmful technologies while framing their work in a morally positive light via connections to hypothetical future benefits. First, we discuss the widespread trend of research in robotics being justified with search-and-rescue applications, the "search" portion of which which may function the same for search-and-surveil or search-and-destroy tasks. Despite a morally positive framing, ostensible robot rescuers often have more obvious applications and funding relationships to systems of oppression like the military industrial complex. Second, we discuss machine learning research, which is widely justified with claims about superhuman abilities to solve our most pressing problems (e.g., the "AI will cure cancer and solve politics" arguments). We argue that it is disingenuous to focus only on (often unrealized) benefits to motivate research that contributes to killing and otherwise harming people, accelerating climate change, and empowering fascist repression. We then present early procedural, technological, and social strategies to mitigate sidepromising in academic publications.

Introduction

The purpose of this WIP research paper is to discuss a pattern that we term "sidepromising." All researchers hope that their work will make the world a better place, but clearly not all technology is exclusively beneficial in its impacts. We define sidepromising as a behavior wherein academics develop harmful technologies while framing their work in a morally positive light via connections to hypothetical future moral benefits and minimizing or ignoring extant and potential harms. Sidepromising leads to papers that are *prima facie* moral by virtue of their promised applications, but whose benefits may not materialize or may be outweighed by unaddressed negatives, thereby creating a clear gap between the stated moral *intent* of researchers and the real-world impacts of their work. We give examples of this behavior from both robotics and machine learning, discuss the harms of sidepromising, and introduce possible steps to mitigate this problem. This work is not an attempt to "call out" specific researchers, but rather to begin challenging the incentive structures that have given rise to this harmful phenomenon. We take inspiration from other critiques of practices in computer science, which present normative frameworks designed to guide technological development toward better outcomes[1,2]. Our work spotlights the role that ostensibly moral academic work plays in enabling the normalized violence of the military-industrial complex, for-profit healthcare industry, and carceral state.

Sidepromising in Robotics Research

"Search and Rescue" Search-and-rescue (SAR) looms large in the minds of robotics researchers, and for good reason; saving lives is a straightforwardly worthwhile application of our work. The phrase "search and rescue" has appeared in 155 publications in the proceedings of the IEEE International Conference on Robotics and Automation (ICRA) alone within the last 20 years[3]. SAR is a hot topic even in social robotics venues. For example, the phrase "search and rescue" has appeared in roughly 192 publications from the ACM/IEEE International Conference on Human-Robot Interaction (HRI) within the last 20 years[4].

Of course, some publications that mention SAR do so only in passing, but many cite SAR as a motivating application for their work. SAR is cited as the motivation for work as diverse as expressing "emotions" in robots through color, sound, and vibration[5] and human-robot swarm interaction for entertainment[6]. Instead of litigating the impossible question of whether all authors who mention SAR truly intend for that to be the primary future application for their research, we ask whether SAR is plausibly among the most likely and most worth discussing applications of the research claiming it as a motivation. In some cases, it is plausible; there absolutely have been a small number of robots designed primarily for SAR. However, much of the robotics work citing SAR as a motivation is either more likely to be used for something else (like search-and-murder or search-and-surveil), or work that does not seem plausibly applicable to SAR. We term these categories "dual-use" and "implausible use" respectively.

Although rare in comparison to the total amount of work claiming to be motivated by SAR, a handful of research efforts have developed specialized SAR robots. For example, the "Robocue"[7] from Kikuchi Manufacturing includes a conveyor belt to move incapacitated and supine/prone humans into the robot's "body" for their protection and transportation. More recently, the ResQbot 2.0[8] appears to use a somewhat similar design without an enclosure around the person being transported. The SemiAutonomous Victim Extraction Robot (SAVER)[9] and Rescue Support Stretcher System[10] are similar. A wheel-legged rescue robot uses a similar stretcher-based transport with different locomotion[11]. The much earlier iRobot Valkyrie Casualty Extraction Concept[12] was just a robot with treads towing a sled that a victim could clamber into. The US Department of Defense later introduced the Robotic Evacuation & Extraction Vehicles (REV/REX) to extract wounded soldiers from the site of injury[13]. The Battlefield Extraction Assist Robot (BEAR)[14,15], the Combat Robotic Nursing Assistant (cRoNA)[16], and the Humanoid Rescue Robot for Calamity Response (HURCULES)[17,18] are semi-humanoid robotic rescue platforms, designed to scoop up and cradle incapacitated humans in their arms.

Dual-Use Rescuers Often robotics research claiming to be motivated by SAR is entirely focused on *search*, with little attention to any *rescue*. We are not the first researchers to notice the relative abundance of research dealing with *search* on its own as opposed to *SAR*[19]. Unfortunately, the search part of SAR can function like the search portion of "search-and-murder" or "search-and-surveil". This research often involves unmanned aerial vehicles (UAVs or "drones")[20]. Drones have been used, with varying degrees of autonomy, in a few hundred SAR operations worldwide[21]. The first reported instance of drone-based autonomous person detection assisting with a rescue was in Poland in 2021[22]. In general, drones with (thermal) cameras could locate victims, establish communications between victims and rescuers, and possibly deliver small amounts of supplies. However, the utility of drones, both potential and realized, for violence and surveillance is also obvious, perhaps more so, and is comparatively sparsely addressed in the robotics literature endowing drones with new capacities. The total number of people *ever* rescued with drone involvement, as reported by a drone manufacturer, is roughly equivalent to the number of people killed in drone strikes just in 2010 in Pakistan alone[21,23].

All kinds of drones are used for killing people, from small 3D printed quadcopters to large multimillion dollar machines[24]. A single Ukranian firm can produce more than 30,000 military drones in a single month, and their operators describe them as "the most important military invention since the bullet."[25]. Drones have fundamentally reshaped war[24]. Since 2001, the US has used drones for (somewhat) targeted killings, a practice that increased in both geography and frequency during the ensuing years[23]. Even commercial drones originally made for nonmilitary purposes like photography have seen large scale systematic use by military actors to drop grenades on civilian populations including children[26]. It is obvious that any drone able to autonomously find and track people will be used for a violent purpose, not (just) for SAR.

Perhaps harmful uses should not prevent us from developing technology that also has benefits, like SAR drones. If we'd had today's drones in 2005, for example, perhaps they could have aided the victims of Hurricane Katrina. This line of reasoning might be more compelling in a sociocultural landscape that prioritizes minimizing harm or promoting wellbeing. Instead, in the actual aftermath of Hurricane Katrina, the mayor of New Orleans ordered police officers to "cease rescue operations and control widespread looting"[27]. In one incident, a group of ununiformed police officers opened fire on unarmed civilians, killing two (one child) and wounding four, and then attempted to cover it up[28]. A society more interested in surveillance and violence than in SAR will use new robotic technologies more for surveillance and violence than for SAR.

Although it is difficult to imagine that roboticists working on drones "for SAR" are unaware of this dual-use problem, they often ignore it and even exacerbate it. For example, the Search and Rescue Image Dataset for Person Detection (SARD)[29] contains (alongside images of possibly injured and exhausted people, which are clearly relevant to SAR) images of people running or appearing to hide underneath cars. With some suspension of disbelief, it is possible to imagine these as relevant to SAR; maybe the person is running to find help for their injured friend, and maybe floodwaters deposited a person and a car in that configuration. However, running or hiding from a drone is also eminently explainable when a person is rightly afraid that the drone will harm them. One might imagine that facilitating automated detection of such people under the sidepromise of SAR could raise a moral issue for some academics.

Drones can clearly be useful for SAR, but roboticists should take this dual-use problem much more seriously than we currently do, and researchers should not be allowed to use this dual-use problem to their advantage. The term "search and rescue" should not function as an incantation to deflect moral criticism from potentially harmful work. We are not the first to identify the dual-use problem in drones. Philippi[30] discusses how modularity contributes to misuse risk in drones and makes this risk especially difficult to mitigate. The mitigation strategies that Philippi[30] delineates are not a complete solution, and are less applicable to SAR drones than to other kinds of drones. They also require government implementation but are unlikely to be supported by many relevant governments. We have deliberately avoided reliance on government in our strategies for mitigating sidepromising as our government in the US becomes increasingly unpredictable.

Implausible Rescuers In many cases, robotics work strains credulity in claiming to be relevant to SAR. This work often involves robots that do not fly. To be credibly useful to SAR, such a robot needs to bring some advantage over a dog, human expert, or drone (we note the development of flying drones for search underground and indoors[20]), while also not obviously better serving some more morally fraught purpose. The "TerminatorBot" is a small (40mm tall) cylindrical robot designed to be deployed via an M203 grenade launcher[31]. To their credit, its creators state that it was designed for surveillance and military reconnaissance alongside, ostensibly, SAR. The Loper robot, presented in a paper titled "A Search and Rescue Robot," is designed to roll up stairs, but has no evident utility to SAR[32,33]. There are countless similar robotics projects claiming to be motivated by SAR. There is no doubt that work like this has value, but there may be doubt about whether that value is primarily relevant to SAR.

Sidepromising in Machine Learning

"Revolutionary AI" in Science and Medicine Numerous works in ML for medicine promise to "revolutionize" healthcare[34] by streamlining parts of the healthcare industry with Large Language Models (LLMs). At least in the US, an LLM tool still gatekept by insurance does not meaningfully increase access to healthcare and largely represents a continuation or degradation of the status quo from the patient's perspective. There is nothing "revolutionary" about technologies that make the current healthcare system more efficient, and thereby more profitable, without improving patient experiences or outcomes. Outside the US, even widespread free access to these models would not be an unmitigated good as LLMs are infamous for being opaque, fallible, unpredictable, and difficult to constrain. Any paper promising medical benefits of LLMs must weigh those against LLMs' proven track record of damaging cognitive functioning[35], promoting dangerous behavior[36], damaging public health via pollutant emissions in proximity to data centers[37], and causing acute individual health problems including new-onset psychosis[38].

Climate Impacts The climate impacts of ML training and inference have been extensively documented[39,40]. Regardless, researchers in ML continue to claim that AI will help humanity tackle climate change. Proposed mechanisms include increasing climate modeling accuracy[41,42,43,44,45,46], engineering improvements to lessen fossil fuel use[47], or climate science communication[48,49]. Improved modeling or reporting are not the same as addressing climate change, and many of these papers do not address the balance between their scientific contributions and their contributions to fossil fuel use and cobalt mining. The research in this space that *does* discuss this balance can serve as a good example that doing so is possible[47,50,51]. Nevertheless, research teams continue to publish papers that emphasize modeling accuracy while neglecting to mention their contribution to the climate crisis. To reiterate, even creating a perfect model of Earth's climate would do nothing by itself to ameliorate climate change, and the field's presumptive techno-solutionist lens hides its contribution to the problem as well as the desperate need for drastic sociocultural action.

Facial Recognition No facial recognition paper published in the 21st century can plausibly claim ignorance of the fact that these technologies are widely used in military and policing contexts[52,53,54,55]. Nevertheless, few papers in the field engage with the moral implications imposed by these use-cases. Papers often present facial recognition as an abstract, politically neutral classification problem[56,57]. It is perhaps understandable that facial recognition researchers would be recalcitrant to say that their classifier will lead to increased targeted violence and state repression, but this is a quintessential example of sidepromising; authors claim increased classification accuracy as an implicit moral good without reckoning with their contribution to harms. Other work examining bias in facial recognition algorithms correctly points out that the algorithms' lower accuracy for darker skinned people can lead to false positives that further erode the civil liberties of already over-policed groups[58]. However, as Raji et al.[59] have argued, facial recognition research can normalize tasks that are inherently harmful to certain communities even when performed 100% accurately. Efforts towards reducing disparities between demographic groups may promote equality, but also improve the tools of the carceral state, which is the militarized enforcement mechanism of the inequalities that capitalism requires. It is impossible to conclusively weigh these concerns against each other, but it is sidepromising to claim improvement on one of them while ignoring the other.

Generation of Harmful Images Given that concerns about fake images pre-date the development of modern generative ML models[60,61,62,63,64], we would expect researchers who create tools like GANs or Latent Diffusion[65] to be aware that these tools can be used for a variety of malicious purposes. For example, currently there are many websites that generate pornographic deepfakes, and it is trivial to download model weights to do so at scale no matter how unconsensual or illegal. Neither the original GAN paper[66] nor a 2020 followup[67] mention social impacts at all; the latent diffusion paper briefly mentions gendered violence towards women but claims that it is "not clear" how much diffusion models contribute to this problem over other kinds of generative models. This omission leads to a diffusion of responsibility whereby early papers fail or neglect to anticipate negative impacts of their work, and later papers view their own contribution as minimal (the cat's out of the bag). Moral considerations are left to follow-up papers written by other researchers[68]. It is not reasonable to expect that authors will anticipate *all* negative applications of their work, but it is obvious that image generation models might naturally be used to generate illicit images that are otherwise difficult to obtain. When autoregressive models are generating pornographic images of children on X.com[69], we must reckon with whether creating that technology carries any culpability and ask - how could this have been avoided?

Causes, Harms, and Mitigation of Sidepromising

Researchers who need to publish their work find themselves pulled in two different directions by two different incentive structures: the peer review process of academic publishing, which values scientific and social progress as defined by academics, and the grant process by which large accumulations of capital (very often the military-industrial complex or healthcare industry) choose which projects receive funding. Researchers need the approval of both their peers and capital. This dynamic produces a pressure to appear moral while serving the amoral motives of profit and empire. There is thus an incentive for researchers to promote framings of their work that allow it to be funded by, e.g., DARPA while still appearing moral to scientific society writ large, including the civilian academics that often evaluate grant proposals (we note that DARPA is not secretive about their intent to use robotic technologies originally developed for SAR towards less morally straightforward purposes[70]). Moreover, grant proposals discuss hypothetical work, and there is an obvious incentive to be generous in imagining its hypothetical benefits.

This pull towards sidepromising begets immoral scientific work by allowing academics to cast ourselves as plausibly moral while developing eventually harmful technologies. Aside from distancing researchers from the real-world harmful effects of their work, it also prevents the research community from seeing imminent future harms. Who better to predict misuse of a technology than those who know it best? I.e., those most incentivized not to predict that.

We want to emphasize that we are not necessarily accusing researchers of being intentionally misleading. Even researchers who we count among the best of humanity are only human, and the human capacity for analgesic self-deception is well documented[71]. "The maintenance of positive self-regard while behaving harmfully is a strong motivator for self-exoneration," and there are a "variety of psychosocial mechanisms by which individuals convince themselves that they are doing good while inflicting harm on others"[72]. Facing the overconstrained demands of chasing funding, publish-or-perish, and one's own conscience is an unenviable position.

Procedural solutions A very simple step for publication venues would be to explicitly instruct that reviewers should not take hypothetical promises in submitted papers at face value. Journals and conferences could also adopt review criteria that take likely future impacts of submissions into account, and give guidelines about how to do so even when impacts are not discussed by the authors. Authors could be provided with examples of future-looking statements that constitute sidepromising, versus ones which are plausible, balanced, and of appropriate scope.

One example is the NeurIPS paper checklist[73,74], which authors complete as a mandatory step of paper acceptance. The checklist explicitly asks authors to "Consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology." This checklist is already addressing some of our sidepromising concerns in spirit. Though not a complete solution, we suggest making these checklist items into review criteria and applying a similar procedure outside of NeurIPS. Another suggestion would be to create an "outcome review" paper category; this type of paper would be similar to a standard review paper, but would be focused on evaluating whether predictions and promises of articles came to pass, as well as their actual impacts. Ethicists and other scholars write *post hoc* reviews[30] covering morally relevant impacts of technological advancements, which may help to retroactively dispel sidepromising, but damage is already done by the time this happens, and these papers may see less penetration into engineering research communities.

Technological Solutions Modern online academic publishing could allow for papers to be dynamic, living documents. One method for encouraging researchers to be honest in the hypothetical future impacts they promise in their paper would be to develop mechanisms which modify the paper if those impacts do not come to pass. Reviewers and authors could identify sections of papers which should be modified or redacted if the authors' claims do not pan out within a certain timeframe. These sections could require authors to submit retroactive citations.

For example, if we say in a paper “this robot will be useful for search and rescue,” we would need to add a citation corroborating that claim within the next eight years or the publisher would redact that claim from the paper or add an author-supplied sentence saying that it did not happen.

Unfortunately, harmful iterations on academic work often take place outside of academia where they are harder to track and sanction. To allow periodic audits for governmental abuse of academic work, authors in the US could submit alongside their papers peer-reviewed Freedom of Information Act (FOIA) requests that could automatically be sent to relevant governmental agencies yearly for some period of time after paper acceptance. Either the authors would need to review the results of those requests, or volunteers as in the peer review process. Online papers could also be accompanied by crowdsourced lists of human rights abuses associated with the technology introduced in the paper. People whose rights are violated obviously lack the incentives towards sidepromising of paper authors. Publishers could provide any of these capabilities as part of their web publishing interface. We envision a future in which a published paper in computer science is accompanied by morally relevant data tracking real-world impacts.

Sociocultural solutions Educators in computer science have a responsibility to address the moral impacts of computational technologies with their students and to help students recognize and read around sidepromising. Teaching about facial recognition or drones without addressing their use cases, or only addressing comfortable use cases, is a kind of pedagogical sidepromising, and it cannot be read as anything other than an attempt to make the future worse.

A solution to sidepromising might start with education, but it cannot end there. Groups of organized academics acting cooperatively have a much better chance of reshaping the incentive structures that constrain their research than individual disorganized researchers. For example, the ‘No Justice, No Robots’ campaign in 2020[75] and the subsequent paper[76] led over 100 roboticists to pledge to “refuse to facilitate the execution or publication of research in robotics or Artificial Intelligence (AI) that is performed in collaboration with State and Local Law Enforcement agencies (or national agencies such as ICE and CBP)” and to “refuse to pursue or accept funding for research projects whose express goal is development of robotic or AI technologies ... intended for use by such agencies.” This campaign shows that there is an appetite for even minimally structured social justice initiatives among researchers, and a campaign explicitly addressing sidepromising could be effective. Finally, we note that the capitalistic incentive structures that determine what science gets done, what technologies get produced, and who gets to use them may seem immutable, but there is no technical reason why this is necessarily the case. Sidepromising could be mitigated as a side effect of broadly antiracist, antifascist, and anticapitalist efforts.

Conclusion

We have described “sidepromising” as the behavior of ignoring a technology’s negative impacts in favor of promising hypothetical beneficial applications. We have explored the causes of this behavior and argued that the field must develop methods for reducing the gap between authorial moral positioning and real-world outcomes. We have proposed some mitigation strategies; more work is required to deploy and evaluate them.

References

[1] Ruha Benjamin. Race after technology. In *Social Theory Re-Wired*, pages 405–415. Routledge, 2023.

[2] Catherine D'ignazio and Lauren F Klein. *Data feminism*. MIT press, 2023.

[3] Search for "search and rescue" in the proceedings of the ieee international conference on robotics and automation (icra). https://ieeexplore.ieee.org/search/searchresult.jsp?queryText=search%20and%20rescue&highlight=true&returnType=SEARCH&matchPubs=true&searchWithin=%22Parent%20Publication%20Number%22:1000639&returnFacets=ALL&ranges=2005_2025_Year. Accessed: 2026-1-18.

[4] Search for "search and rescue" in publications from the acm/ieee international conference on human-robot interaction (hri). https://dl.acm.org/action/doSearch?AllField=%22search+and+rescue%22&ConceptID=119235. Accessed: 2026-1-18.

[5] Sichao Song and Seiji Yamada. Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 2–11, 2017.

[6] Javier Alonso-Mora, Roland Siegwart, and Paul Beardsley. Human-robot swarm interaction for entertainment: From animation display to gesture based control. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 98–98, 2014.

[7] Kazuyoshi Miyazawa. Fire robots developed by the Tokyo Fire Department. *Advanced Robotics*, 16(6): 553–556, 2002.

[8] Roni Permana Saputra, Nemanja Rakicevic, Isabelle Kuder, Joel Bilsdorfer, Alexander Gough, Alexandra Dakin, Emma de Cocker, Shaun Rock, Richard Harpin, and Petar Kormushev. Resqbot 2.0: An improved design of a mobile rescue robot with an inflatable neck securing device for safe casualty extraction. *Applied Sciences*, 11(12):5414, 2021.

[9] P Ben-Tzvi, A Williams, B Sebastian, A Kumar, and W Saab. Semi-autonomous victim extraction robot (SAVER). *US Provisional Patent Application*, (62/660,869), 2018.

[10] Yuki Iwano, Koichi Osuka, and Hisanori Amano. Development of rescue support stretcher system with stair-climbing. In *2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*, pages 245–250. IEEE, 2011.

[11] Meng Ning, Zefeng Ma, Hanlong Chen, Jun Cao, Changhong Zhu, Yifan Liu, and Yuefeng Wang. Design and analysis for a multifunctional rescue robot with four-bar wheel-legged structure. *Advances in Mechanical Engineering*, 10(2):1687814017747399, 2018.

[12] Gary Gilbert, Troy Turner, and Ron Marchessault. Army medical robotics research. 2007.

[13] Gary R Gilbert and Michael K Beebe. United states department of defense research in robotic unmanned systems for combat casualty care. Technical report, 2010.

[14] Andrew C Yoo, Gary R Gilbert, and Timothy J Broderick. Military robotic combat casualty extraction and care. In *Surgical Robotics: Systems Applications and Visions*, pages 13–32. Springer, 2010.

[15] Daniel Theobald. Mobile extraction-assist robot, May 18 2010. US Patent 7,719,222.

[16] John Hu and Yi-Je Lim. Robotic first responder system and method, June 5 2014. US Patent App. 13/691,840.

[17] Byunghun Choi, Wonsuk Lee, Gyuhyun Park, Youngwoo Lee, Jihong Min, and Seongil Hong. Development and control of a military rescue robot for casualty extraction task. *Journal of Field Robotics*, 36(4):656–676, 2019.

[18] Byunghun Choi, Gyuhyun Park, and Youngwoo Lee. Practical control of a rescue robot while maneuvering on uneven terrain. *Journal of Mechanical Science and Technology*, 32(5):2021–2028, 2018.

[19] Roni Permana Saputra and Petar Kormushev. Resqbot: A mobile rescue robot for casualty extraction. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 239–240, 2018.

[20] Mingyang Lyu, Yibo Zhao, Chao Huang, and Hailong Huang. Unmanned aerial vehicles for search and rescue: A survey. *Remote Sensing*, 15(13):3266, 2023.

[21] Drone rescues around the world. `https://enterprise.dji.com/drone-rescue-map/`. Accessed: 2026-1-12.

[22] Tomasz Niedzielski, Mirosława Jurecka, Bartłomiej Miziński, Wojciech Pawul, and Tomasz Motyl. First successful rescue of a lost person using the human detection system: A case study from Beskid Niski (SE Poland). *Remote Sensing*, 13(23):4903, 2021.

[23] Sarah Kreps and John Kaag. The use of unmanned aerial vehicles in contemporary conflict: A legal and ethical analysis. *Polity*, 44(2):260–285, 2012.

[24] Erik A Davis. Drones and the changing character of war. *The US Army War College Quarterly: Parameters*, 55 (4):7, 2025.

[25] The Wall Street Journal. On the front lines with Ukraine's killer drone pilot, accessed January 12, 2026. URL `https://www.youtube.com/watch?v=062R1k54Ijo`.

[26] Yuval Abraham. 'Like a video game': Israel enforcing Gaza evacuations with grenade-firing drones: The Israeli army is weaponizing Chinese-made drones to police expulsion orders across Gaza, with soldiers saying they deliberately target civilians so others will 'learn' not to return, an investigation reveals. `https://www.972mag.com/drones-grenades-gaza-chinese-autel/`. Accessed: 2026-1-12.

[27] Dan Berger. Constructing crime, framing disaster: Routines of criminalization and crisis in Hurricane Katrina. *Punishment & society*, 11(4):491–510, 2009.

[28] Jarrett M Drake. Insurgent citizens: the manufacture of police records in post-Katrina New Orleans and its implications for human rights. *Archival Science*, 14(3):365–380, 2014.

[29] Sasa Sambolek and Marina Ivasic-Kos. Automatic person detection in search and rescue operations using deep CNN detectors. *Ieee Access*, 9:37905–37922, 2021.

[30] Martina Philippi. The ambivalence of multi-purpose design: On the dual-use and misuse risks of humanitarian UAVs. *Research Ethics*, page 17470161251344321, 2025.

[31] Richard M Voyles. Terminatorbot: a robot with dual-use arms for manipulation and locomotion. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, volume 1, pages 61–66. IEEE, 2000.

[32] Sam Herbert, Nathaniel Bird, Andrew Drenner, and Nikolaos Papanikolopoulos. A search and rescue robot. In *2009 IEEE International Conference on Robotics and Automation*, pages 1585–1586. IEEE, 2009.

[33] Sam D Herbert, Andrew Drenner, and Nikolaos Papanikolopoulos. Loper: A quadruped-hybrid stair climbing robot. In *2008 IEEE International Conference on Robotics and Automation*, pages 799–804. IEEE, 2008.

[34] Shuroug A Alowais, Sahar S Alghamdi, Nada Alsuhebany, Tariq Alqahtani, Abdulrahman I Alshaya, Sumaya N Almohareb, Atheer Aldairem, Mohammed Alrashed, Khalid Bin Saleh, Hisham A Badreldin, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*, 23 (1):689, 2023.

[35] Hao-Ping Lee, Advait Sarkar, Lev Tankelevitch, Ian Drosos, Sean Rintel, Richard Banks, and Nicholas Wilson. The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence

effects from a survey of knowledge workers. In *Proceedings of the 2025 CHI conference on human factors in computing systems*, pages 1–22, 2025.

[36] Eileen Guo. An AI chatbot told a user how to kill himself—but the company doesn't want to "censor" it. `https://www.technologyreview.com/2025/02/06/1111077/nomi-ai-chatbot-told-user-to-kill-himself/`, 2025. Accessed: 2026-1-18.

[37] Yuelin Han, Zhifeng Wu, Pengfei Li, Adam Wierman, and Shaolei Ren. The unpaid toll: Quantifying the public health impact of AI. *arXiv preprint arXiv:2412.06288*, 2024.

[38] Joseph M. Pierre, Ben Gaeta, Govind Raghavan, and Karthik V. Sarma. "You're not crazy": A case of new-onset AI-associated psychosis. *Innovations in Clinical Neuroscience*, 22(10-12), 2025.

[39] Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research*, 21(248):1–43, 2020.

[40] Qiang Wang, Yuanfan Li, and Rongrong Li. Ecological footprints, carbon emissions, and energy transitions: the impact of artificial intelligence (AI). *Humanities and Social Sciences Communications*, 11(1):1–18, 2024.

[41] Chris Huntingford, Elizabeth S Jeffers, Michael B Bonsall, Hannah M Christensen, Thomas Lees, and Hui Yang. Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters*, 14(12):124007, 2019.

[42] Andrew Crane-Droesch. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11):114003, 2018.

[43] Paul A O'Gorman and John G Dwyer. Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, 10 (10):2548–2563, 2018.

[44] Sina Ardabili, Amir Mosavi, Majid Dehghani, and Annamária R Várkonyi-Kóczy. Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review. In *International conference on global research and education*, pages 52–62. Springer, 2019.

[45] Cristian Bodnar, Wessel P Bruinsma, Ana Lucic, Megan Stanley, Anna Allen, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan A Weyn, Haiyu Dong, et al. A foundation model for the Earth system. *Nature*, pages 1–8, 2025.

[46] Veronika Eyring, William D Collins, Pierre Gentine, Elizabeth A Barnes, Marcelo Barreiro, Tom Beucler, Marc Bocquet, Christopher S Bretherton, Hannah M Christensen, Katherine Dagon, et al. Pushing the frontiers in climate modelling and analysis with machine learning. *Nature Climate Change*, 14(9):916–928, 2024.

[47] David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)*, 55(2):1–96, 2022.

[48] Carmen Atkins, Gina Girgente, Manoochehr Shirzaei, and Junghwan Kim. Generative AI tools can enhance climate literacy but must be checked for biases and inaccuracies. *Communications Earth & Environment*, 5(1): 226, 2024.

[49] David Thulke, Yingbo Gao, Petrus Pelser, Rein Brune, Rricha Jalota, Floris Fok, Michael Ramos, Ian Van Wyk, Abdallah Nasir, Hayden Goldstein, et al. ClimateGPT: Towards AI synthesizing interdisciplinary research on climate change. *arXiv preprint arXiv:2401.09646*, 2024.

[50] Josh Cowls, Andreas Tsamados, Mariarosaria Taddeo, and Luciano Floridi. The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations. *Ai & Society*, 38(1): 283–307, 2023.

[51] Neeta Kumari and Soumya Pandey. Application of artificial intelligence in environmental sustainability and climate change. In *Visualization techniques for climate change with machine learning and artificial intelligence*, pages 293–316. Elsevier, 2023.

[52] Isadora Neroni Rezende. Facial recognition in police hands: Assessing the 'clearview case' from a European perspective. *New Journal of European Criminal Law*, 11(3):375–389, 2020.

[53] Pete Fussey, Bethan Davies, and Martin Innes. 'Assisted' facial recognition and the reinvention of suspicion and discretion in digital policing. *The British journal of criminology*, 61(2):325–344, 2021.

[54] Forrest E Morgan, Benjamin Boudreaux, Andrew J Lohn, Mark Ashby, Christian Curriden, Kelly Klima, and Derek Grossman. Military applications of artificial intelligence: ethical concerns in an uncertain world. Technical report, 2020.

[55] Christopher S Milligan. Facial recognition technology, video surveillance, and privacy. *S. Cal. Interdisc. LJ*, 9: 295, 1999.

[56] Wenyi Zhao, Rama Chellappa, P Jonathon Phillips, and Azriel Rosenfeld. Face recognition: A literature survey. *ACM computing surveys (CSUR)*, 35(4):399–458, 2003.

[57] Mei Wang and Weihong Deng. Deep face recognition: A survey. *Neurocomputing*, 429:215–244, 2021.

[58] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR, 2018.

[59] Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Remi Denton. Saving face: Investigating the ethical concerns of facial recognition auditing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 145–151, 2020.

[60] Thomas H Wheeler. *Phototruth or photofiction?: Ethics and media imagery in the digital age*. Routledge, 2005.

[61] Rose Wiles, Jon Prosser, Anna Bagnoli, Andrew Clark, Katherine Davies, Sally Holland, and Emma Renold. Visual ethics: Ethical issues in visual research. 2008.

[62] Scott R Stroud. The dark side of the online self: A pragmatist critique of the growing plague of revenge porn. *Journal of Mass Media Ethics*, 29(3):168–183, 2014.

[63] Carolyn A Uhl, Katlin J Rhyner, Cheryl A Terrance, and Noël R Lugo. An examination of nonconsensual pornography websites. *Feminism & Psychology*, 28(1):50–68, 2018.

[64] Rosemary Ricciardelli and Michael Adorjan. 'If a girl's photo gets sent around, that's a way bigger deal than if a guy's photo gets sent around': Gender, sexting, and the teenage years. *Journal of gender studies*, 28(5): 563–577, 2019.

[65] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[66] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[67] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144, 2020.

[68] David Gray Widder, Dawn Nafus, Laura Dabbish, and James Herbsleb. Limits and possibilities for "ethical AI" in open source: A study of deepfakes. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 2035–2046, 2022.

[69] AI tool Grok used to create child sexual abuse imagery, watchdog says. www.theguardian.com/technology/2026/jan/08/ai-chatbot-grok-used-to-create-child-sexual-abuse-imagery-watchdog-says. Accessed: 2026-1-12.

[70] Gill A Pratt. Robot to the rescue. *Bulletin of the Atomic Scientists*, 70(1):63–69, 2014.

[71] William Von Hippel and Robert Trivers. The evolution and psychology of self-deception. *Behavioral and brain sciences*, 34(1):1–16, 2011.

[72] Albert Bandura. Self-deception: a paradox revisited. *Behavioral & Brain Sciences*, 34(1), 2011.

[73] Joelle Pineau, Philippe Vincent-Lamarre, Koustuv Sinha, Vincent Larivière, Alina Beygelzimer, Florence d'Alché Buc, Emily Fox, and Hugo Larochelle. Improving reproducibility in machine learning research (a report from the NeurIPS 2019 reproducibility program). *Journal of machine learning research*, 22(164):1–20, 2021.

[74] NeurIPS paper checklist. https://neurips.cc/public/guides/PaperChecklist. Accessed: 2026-1-10.

[75] No justice, no robots: An open letter from robotics researchers. https://nojusticenorobots.github.io/. Accessed: 2026-1-18.

[76] Tom Williams and Kerstin Sophie Haring. No justice, no robots: From the dispositions of policing to an abolitionist robotics. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pages 566–575, 2023.