# Physicists Must Engage with AI Ethics, Now

Physicists are increasingly utilizing AI and even driving its development, but we cannot divorce ourselves from the ethical implications and impacts of this technology.

By **Savannah Thais**

Popular media depictions of AI often involve apocalyptic visions of killer robots, humans mined for resources, or the elimination of the human race altogether. Even the rosier visions of an AI-driven world imagine most traditional human efforts and behaviors replaced with machines. Our collective imaginations of AI are often focused on this "singularity"—an irreversible point when technology overtakes humanity.

However, the realization of this kind of artificial general intelligence (AGI), where a machine can perform any cognitive task a human can, remains a long way off. It is true that there have been impressive advances in machine learning (ML) and that algorithms can now best humans at various tasks. Yet, in many ways, we are not much closer to achieving AGI than we were in 1951 when pioneering computer science researcher **Alan Turing predicted** that machines would "outstrip our feeble powers" and "take control."
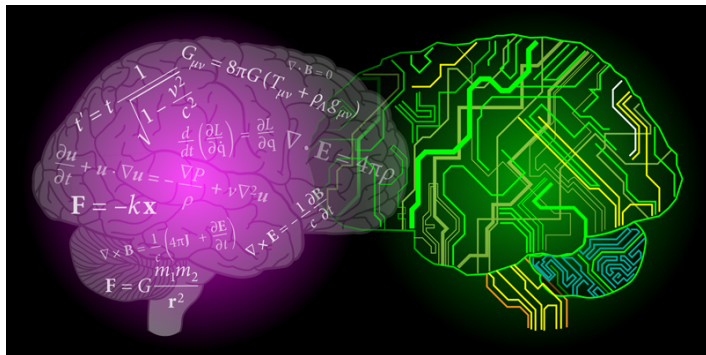
Credit: APS/**Alan Stonebraker**

While it's important to think about the ethical implications of AGI, preoccupation with the singularity often eclipses the fact that we are already living in a technological dystopia. Governments and organizations worldwide already rely on uninterpretable, under-tested, third-party algorithms to make critical decisions around **bail setting**, **employment**, **healthcare access**, and more. **Private corporations**, **governments**, and even individuals are deploying facial recognition and other automated surveillance technologies to **track private citizens** and suppress **minority groups** (see also **QA: When Politics and Particles Collide**). The rare audits of these algorithms often reveal that their predictions show biases based on **race**, **gender**, and **other factors** or are simply inaccurate.

These harmful impacts of AI have occurred, in part, because the rate of innovation has far outpaced the development of legislative oversight and because technology companies have been reluctant to (or have refused to) self-regulate. However, change may be coming. Activists and researchers are increasingly calling for governments to **limit or outright ban** certain technologies, many countries are developing **standards** for AI deployment, and technology companies large and small are **creating or growing their ethics and fairness teams**. Since I started attending AI conferences in 2016, I've seen the number of workshops, papers, and conversations around ethics in current AI grow. This year, the Neural Information Processing Systems meeting (NeurIPS)—one of the world's largest AI research conferences—will require all submissions to include an **impact statement** discussing "ethical aspects and future societal consequences" and acknowledging potential conflicts of interest.

So why am I writing about this in a physics publication? Well, it's no secret that AI and ML are increasingly a core component of physics research programs. About half of the **pre-conference courses and tutorials** planned for this year's March Meeting of the American Physical Society were focused on AI. Additionally, many workshops, lecture series, and summer schools deal exclusively with the intersection of these fields.

However, physicists aren't just recycling existing ML methods and applying them to physics questions. Many physics graduate students and postdocs leave the field to become AI researchers or engineers in tech companies and research institutes. Other physicists undertake cross-disciplinary research that advances both fields; some even hold joint appointments across physics and computer science. In fact, the US National Science Foundation included "AI for Discovery in Physics" as one of six themes in its call for cross-disciplinary **National AI Research Institutes**.

Physicists have the opportunity to play a critical role in understanding and regulating AI. Physics research often requires a careful analysis of algorithmic bias or interpretability, which are critical methods for improving the fairness of AI systems. We are becoming increasingly impactful in foundational AI research spaces and have opportunities to effect real change. We have the technical knowledge to help make AI understanding accessible to the public. Moreover, as community members, we have a duty to participate in these conversations and efforts. We cannot separate ourselves from the global ethical implications of the algorithms we build simply because our primary intent is to use them to enable physics research.

If you're grappling with AI ethics for the first time (or even if you've thought about these issues many times before), I encourage you to start with readings from prominent scholars in the field. **Stanford's Institute for Human-Centered Artificial Intelligence**, **the AI Now Institute at NYU**, Harvard's **Berkman Klein Center**, Oxford's **Future of Humanity Institute**, and the **Alan Turing Institute** publish excellent research and host accessible, interdisciplinary conversations and events. If you can attend AI conferences like NeurIPS or the International Conference on Learning Representations (ICLR), participate in sessions on AI ethics, data rights, and interpretability. Learn about ways AI is used in your local community and about the effects it's having on those around you, particularly on vulnerable populations.

Then, make your voice heard. Contribute to discussions at your institute or at conferences. Look for instances where research could have questionable ethical implications and push back by starting conversations with the authors and research community. Encourage your colleagues, students, and collaborators to learn about how these technologies are used beyond the lab and advocate to policy makers when these uses are harmful. Offer your technical skills to support efforts against prejudiced applications of AI or unwarranted surveillance at a community level. Promote interdisciplinary research and use your position to elevate other voices. Equip people in your communities with information about how AI impacts their lives and what actions they can take to combat unethical implementations of AI.

AI is shaping our world, our lives, our rights, and our futures every day. As scientists and citizens, we must actively engage to ensure AI develops and is utilized in an ethical and equitable manner.

Savannah Thais: Princeton Institute for Computational Science and Engineering, Princeton University, Princeton, New Jersey, USA