AI and Ethics (2021) 1:297–300
https://doi.org/10.1007/s43681-020-00031-2

ORIGINAL RESEARCH

Opening the path to ethics in artificial intelligence

Kelly Forbes[1]

Received: 16 September 2020 / Accepted: 26 November 2020 / Published online: 4 February 2021

Abstract
In an attempt to address the ethical challenges in AI, we currently have several ethical AI frameworks in place worldwide, with more being released or in development around the world every day. From the European Commission's Guidelines for Trustworthy AI, to the Asilomar AI Principles, the message is usually similar: more transparency and explicability. However, navigating the broad number of resources currently available is not a simple process. So, how do we find common ground when it comes to ethics in AI?

Keywords Trustworthy AI · Ethical AI · Artificial Intelligence

In many respects, artificial intelligence (AI) is still in its youth and recent advancements have only been made possible due to the vast increases in the volume of data. Although slow in comparison to some of the earliest predictions, the field of AI has recently made striking advances. Many experts argue that it is likely that we will see significant breakthroughs sometime in this century, possibly reaching an artificial general intelligence level [1]. Others believe we are still far from this ideal:

> Getting to that level-general-purpose artificial intelligence with the flexibility of human intelligence isn't some small step from where we are now; instead it will require an immense amount of foundational progress—not just more of the same sort of thing that's been accomplished in the last few years, but—as we will show—something entirely different [2].

While it's been said that, when it comes to AI, we have only scratched the surface so far, we can also see how these relative small developments have already significantly echoed drastic changes—social, ethical and political.

The ethical debate is not new [3]; however, as these developments unfold at a faster pace, the time calls for more concrete discussions around the ethics of AI, at a global level.

In an attempt to address some of these challenges, we currently have several ethical AI frameworks in place worldwide, with more being released or in development around the world every day. From the European Commission's Guidelines for Trustworthy AI, to the Asilomar AI Principles, the message is usually similar: more transparency and explicability. However, navigating the broad number of resources currently available is not a simple process. More importantly, when it comes to building AI, we are far from the practice of these ideals. There is still no common or unifying discussion on how to govern ethics in AI implementations, or the ongoing auditability once machine learning is improving without human intervention.

In a recent paper [4], the author refers to this problem as "principle proliferation". There are too many different frameworks available, not only making it difficult to navigate through them but also opening an opportunity for choice when there should not be one.

Additionally, in a recent study [5], researchers found that the effectiveness of ethical guidelines or ethical codes is almost zero and that they do not change the behavior of professionals from the tech community.

More importantly:

> Ethical research also requires internalizing a commitment to it, aided by training and education on codes and appropriate research methods, mentoring and workplace cultures that foster ethics, transparency about how the research was conducted, and forums (in person and in writing, local and international) where researchers can share their experiences and the challenges they face." [6]

✉ Kelly Forbes
kelly@aiasiapacific.org

[1] AI Asia Pacific Institute, Singapore, Singapore

As we reflect on the above findings, we can conclude that not much progress will be made until a clear plan to practice these ethical principles is drawn. I envisage that this process would have to comprise three key stages:

1 Uniformity

A global agreement on what these ethical values are is critical. Artificial intelligence is not contained by borders. The only way to exercise a concise approach is through investing in international collaboration. We need a practical engagement framework which surpasses territory to cater for the impact of AI applications.

In 1948, the United Nations released the Universal Declaration of Human Rights, which sets out the basic rights and freedoms that apply to all men, women and children, regardless of background, place of residence, appearance and beliefs. It was the first international agreement on the basic principles of human rights. When it comes to the foundational principles of artificial intelligence we need a comparable agreement.

Recently, the United Nations Secretary-General released a roadmap for digital cooperation which envisages eight key areas for action. Out of these eight key areas, supporting global cooperation on artificial intelligence is one of them. This is a critical step in the right direction.

Moving towards an actionable approach, however, will require much collaboration between stakeholders across government, business, academia and broader society to better navigate the challenges ahead. We should aim to design ethical guidelines that can be easily understood and are able to be carried through, despite territorial boundaries.

2 Education

This is the process of educating society and the AI community as to the principles encouraging the development of ethical AI. This is a crucial and long step, essentially moving from theory to practice. In fact, even if we draw a perfect ethical framework, how to adhere to this is the barrier in our path to ethical AI.

In practice, organisations are still struggling to make sense of this evolving industry. There are several ethical developments, often distinct from country to country, but for international organisations which operate internationally, the compliance process can be tedious.

During this education process, much awareness needs to be developed around why ethics in AI is important. But beyond this, the development of a process that can simplify these requirements in the industry is equally important. The considerations of responsible AI are not only for developers and practitioners but, in moving forward, are needed in every industry. The investment industry, for example, will play a major role in this step:

2.1 Investment

The increase in investment opportunities in AI has been one of the pillars in its development, but it also creates new vulnerabilities. Ethics in AI is an emerging yet critical consideration in responsible investment [7]. We are daily witnessing companies being fined billions for non-compliance with ethical and data issues. Investors will now be more often required to understand the risks involved in each AI solution or company that they choose to invest in. This is no longer simply an option; understanding the minimum technical, ethical and regulatory consequences of these investments is quickly becoming a requirement.

If regulatory consequences are not sufficient to encourage investors to start thinking about ethics in AI, consider this:

> Three-fourths of consumers today say they won't buy from unethical companies, while 86% say they're more loyal to ethical companies, according to the 2019 Edelman Trust Barometer. In Salesforce's recent Ethical Leadership and Business survey, 93% of consumers say companies have a responsibility to positively impact society. Businesses are being held more accountable than ever for what they do and how they behave. [8]

These concerns about ethics are demanding improvements in how AI companies build solutions and should create more and more interest for investors to demand the right approach from their investments [9].

In a recent conversation with Janet Wong, CFA from the EOS at Federated Hermes Asia and global emerging markets stewardship team, she examined these rising expectations from investors. Investors are increasingly expecting to see the following requirements when choosing AI companies to invest in:

(a) evidence of AI governance and oversight within the company, including clear responsibility on the board level to oversee AI-related issues;
(b) evidence of public commitment to trust the AI; and
(c) evidence of how the company is operationalising these ethical principles.

Investments also open for a great opportunity to shape the market. While many founders intuitively start their journey with good intentions for their technology, along the way,

many will eventually cross a path where keeping true to their values will become excruciatingly challenging.

Rana el Kaliouby is the founder of Affectiva, a startup that is developing Emotion AI that can detect emotion just the way humans do, from multiple channels. The technology can be applied to help children with Asperger syndrome read and respond to facial expressions. Rana called it the "emotional hearing aid". She recently shared her challenge in keeping true to this purpose and being able to raise funds, during a podcast conversation. For many founders, this means rejecting proposals and offers from government or other organisations where their technology might be distorted; in her case this could be an excellent tool for surveillance, if in the wrong hands. The information generated by such technology could be used in various harmful ways, from surveillance by the government, to impacting the likelihood of people from a certain race getting jobs. In fact, there are already companies selling predictions for how likely someone is to become a terrorist or paedophile.

In such cases, founders are fully aware of the dangers of the technology they are creating being misused. Most have a full grasp of the trustworthy AI principles, and they are true to their purpose. Unfortunately, some will lose momentum—not due to a change of heart, but because it's too difficult for startups to survive in the current climate.

James Brusseau has recently proposed an ethical evaluation of AI-intensive companies which might allow investors to knowledgeably participate in the decision [10]. He argues that artificial intelligence, like other contemporary technologies, should go through the following categories of evaluation: autonomy, dignity, privacy and performance. Combined, these categories would form a robust and credible model for humanitarian investing in AI-intensive companies.

If the market encourages companies to change their ideals to fulfil economic advantages, it is very difficult for us to navigate this. In a practical way, it is only by addressing these core foundations that we can really impose improvements. One way to do that is through the development of a more mindful investment system. As more investors demand ethics in AI, the industry might be pushed to embrace these principles and to learn, how to live them.

3 Accountability

Finally, society has shown that not much adherence ever occurs, unless we have a clear system of accountability in place. Regulation might not necessarily be the right response; in fact, concerns that poorly designed regulation could slow down innovation are well sustained. As suggested by Roger Clarke [11], an approach which considers regulatory alternatives for AI—such as self and industry regulation, co-regulatory arrangements and formal law—should be carefully designed and evaluated given the technical and political complexities.

Some of these suggested accountability mechanisms are surging in the industry to encourage the adoption of AI. There is a growing preference towards engaging AI companies which can demonstrate an internal commitment to ethics in AI, for example, if they have adopted an internal policy overseeing these ethical developments. This enables trust, which in exchange facilitates innovation. As mentioned above, consumers are increasingly unlikely to engage businesses with unethical companies.

An accountability system would encourage that data security, non-bias and transparency are enhanced when designing AI. I envisage a near future where ethical AI is possible and made accessible to the world; where these principles are the living foundation of every AI development, encouraging the development and progress of trustworthy AI.

4 Conclusion

In seeking a common ground when it comes to ethics in AI, the steps mentioned above are crucial initial steps to enable change beyond mere theoretical frameworks. First, it is important that we arrive at some type of consensus when it comes to ethical principles across different territories. Then, we are required to pass this message across and build a community—a step that we understand as education. Perhaps it is through education that we might encourage the development of a more mindful investment system, which encourages the growth of companies that reflect these ideals. Finally, without some type of accountability system, it will be difficult to move from theory to practice. More importantly, cooperation is what will help us to navigate this process.

Funding Not applicable.

Availability of data and material All material has been included under reference section.

Compliance with ethical standards

Conflict of interest Not applicable.

Code availability Not applicable.

References

1. Bostrom, N.: Superintelligence: paths, dangers, strategies. Oxford University Press, Oxford (2014)

2. Marcus, G., Davis, E.: Rebooting AI: building artificial intelligence we can trust. Pantheon Books, New York (2019)
3. Samuel, A.L.: Some moral and technical consequences of automation—a refutation. Science (1960). https://doi.org/10.1126/science.132.3429.741
4. Floridi, L., Cowls, J.: A unified framework of five principles for AI in society. HDSR (2019). https://doi.org/10.1162/99608f92.8cd550d1
5. Weinbaum, C., Landree, E., Blumenthal, M. S., Piquado, T., Gutierrez, C. I.: Ethics in scientific research. RAND Corporation (2019)
6. Weinbaum et al. (n 9)
7. Christine, C., Katherine, F., Sonya, L., Nicholas, S., Janet, W.: Investors' expectations on responsible artificial intelligence and data governance. Hermes (2019)
8. Richard S.: Why ethical AI is a critical differentiator. Forbes (2019)
9. Jared Council: Investors urge AI startups to inject early dose of ethics. Wall Street J. (2019)
10. Brusseau, J.: AI human impact toward a model for ethical investing in AI-intensive companies. (2020). https://doi.org/10.13140/RG.2.2.33213.28644
11. Roger, C.: Regulatory alternatives for AI. ScienceDirect (2019). http://www.rogerclarke.com/EC/AIR-Final.pdf