(CT5151) Privacy for AI Professionals

Alexey Shapovalov (20235952)

January 9, 2022

1 Introduction

As modern society progresses the use of artificial intelligence (AI) is much more prevalent across many businesses and sectors. Technological advancements such as high speed networks and large data centers facilitate the collection of a large quantities of rich data points used to fuel AI systems. As this trend continues, privacy becomes increasingly more important as the vast amount of personal information collected can easily be misused resulting in breaches of privacy. AI professionals need to proactively consider privacy throughout their work to ensure privacy invasions do not occur. This essay critically reflects on privacy concerns and approaches for AI professionals.

2 Understanding Privacy

There are many reasons as to why privacy is important but at its essence, privacy is a basic human right:

> *"Privacy is the right to be let alone."* (Warren and Brandeis, 1890)

The most popular counter-argument to privacy is that if an individual only engages in legal activity, they have nothing to hide. Therefore there is no harm in collecting their information and in fact they should not really care. However, Solove (2007) shows how this argument and its common variants are flawed.

This is because they are based on a view that privacy is solely for the purpose of concealment or secrecy. A simple, yet compelling illustration is that if people have nothing to hide, then a stranger should be more than happy to oblige being photographed naked. This is obviously not the case, even if the stranger does not engage in incriminating behaviour and has "nothing to hide".

In the modern world, the concept of what specifically is privacy is extremely elusive – definitions are either too broad or too narrow to sufficiently encompass the topic. As the power of technology advances and the envelope of what is possible expands, understanding privacy becomes even more challenging. Solove (2006) defines a taxonomy to privacy which consists of four general categories (and sixteen different subcategories):

1. Information Collection – the collection aspect of privacy. What information is collected? Is it collected legitimately?

2. Information Processing – the data processing aspect of privacy. How is the data stored and manipulated?

3. Information Dissemination – the data access aspect of privacy. What parties have access to the data? Is the data distributed to unknown parties?

4. Invasion – the invasive and intrusive aspect of privacy.

This taxonomy gives AI professionals a grounding in terms of evaluating if their work breaches privacy. A given AI system can be viewed in terms of these four categories to understand the privacy implications of the system.

There are countless of sensational news produced as a consequence of privacy invasions in AI. The most well known is the Cambridge Analytica scandal where the data of 87 million Facebook users was used, without consent, for the purposes of political advertising (Confessore, 2018). Such a scandal is a clear violation of privacy as the data was collected through a psychology questionnaire that required access to a user's Facebook account. Another example is when Amazon attempted to create an AI to aid in the recruitment process, a system that given some resumes could select the top candidates (Dastin, 2018). Due to algorithmic

bias (reflective of the gender bias in the tech industry) the AI favoured male candidates and needed to be scraped. If data is stored in an insecure manner personal information can easily end up in the hands of criminals. IBM estimates the average cost of a data breach in 2021 is 4.24 million USD (IBM, 2021).

3 Privacy & AI Professionals

AI professionals are usually thought of as the individuals who build AI models. In reality many cross functional disciplines need to be involved for a successful AI deployment. This includes roles such as data engineers, data scientists, managers, architects, developers and more. Furthermore, organisations are typically structured in a hierarchical manner, engineer, manager, executive. As such an AI professional can take many forms and each should be aware of privacy concerns. The Ethics Guidelines for Trustworthy AI (European-Commission, 2019) is an excellent resource that all AI professionals should be aware of. It outlines both technical and non-technical methods on implementing trustworthy AI. The guide puts forth seven requirements of trustworthy AI that can be used to assess systems:

1. Human agency and oversight – AI systems should "support human autonomy and decision-making."

2. Technical robustness and safety – AI systems should be technically robust in terms of how they behave. Their behaviour should be evaluated to work as expected.

3. Privacy and data governance – data used by AI systems should be evaluated for quality and integrity.

4. Transparency – AI systems should be traceable, explainable and it should be known that an AI is not human when interacting with users.

5. Diversity, non-discrimination and fairness – diversity and inclusion should be considered in the creation of AI.

6. Societal and environmental wellbeing – AI should not be detrimental, it should benefit all human beings, including future generations.

7. Accountability – Mechanisms should be put in place to ensure responsibility and accountability of AI systems.

In addition to this, the Privacy by Design framework (Cavoukian, 2009) was designed to aid AI professionals in ensuring a high privacy standard for AI systems. It consists of seven principles that form a "design-thinking" perspective to incorporating privacy into a system:

1. Proactive not Reactive; Preventative not Remedial – privacy should be approached proactively rather than reactively. AI professionals should anticipate privacy invasions rather than react to them.

2. Privacy as the Default Setting – privacy should be the default. "If an individual does nothing, their privacy still remains intact."

3. Privacy Embedded into Design – privacy should be integral to the system. It should be considered core functionality.

4. Full Functionality: Positive-Sum, not Zero-Sum – there should not be a trade-off in achieving privacy.

5. End-to-End Security: Full Lifecycle Protection – the entire lifecycle of data collected should be secure.

6. Visibility and Transparency: Keep it Open – business practises and technologies involved should always be "operating according to the stated promises and objectives."

7. Respect for User Privacy: Keep it User-Centric – above all the interests of the individual should be considered with the utmost of importance.

Much like there are many approaches and techniques to building robust software, the Privacy by Design framework provides AI professionals with an approach to ensuring the privacy of the end user is thought through in advance.

4 Privacy Policies

In today's world privacy is largely dealt with in a "notice-and-consent" manner via long and often highly technical documents known as privacy policies. However, there are a number of issues associated with this approach and AI professionals should strive to do better (Nissenbaum, 2011). Notice-and-consent treats privacy in a "take it or leave it" fashion and is usually implemented in an "opt out" fashion (see the second and fourth principle in Section 3). Unfortunately the cost of opting out is often too high a price to pay and as a consequence individuals agree to privacy policies without really seeing it as a choice. For example refusing to use Facebook because of a concerning privacy policy may not be an option for a person who's main means of communication is the Facebook messenger app. Another issue is the "transparency paradox", it is impossible to textually define how your data will be used without the text being too difficult (or too timely) to comprehend and analyze. AI is an extremely complex field often requiring specialised degrees to participate, it is infeasible to expect the average person to truly understand how their data will be used for the purposes of AI. Furthermore most people use many different services which all have their own privacy policies, is it really fair to expect that each policy is thoroughly read prior to use of service?

5 Contextual Integrity

Nissenbaum (2004) puts forth an alternative approach to notice-and-consent. Instead privacy is viewed in terms of "context-relative norms" of which the key parameters "are actors (subject, sender, recipient), attributes (types of information), and transmission principles (constraints under which information flows)". The approach is captivatingly simple, information flow should abide by the entrenched norms under the context of which it is collected. For example, in the case of healthcare; the entrenched norm is that when a patient shares information with a doctor, the expectation is that the doctor may share it with

necessary specialists but certainly will not share the information with advertisement agencies. A clear cut illustration of where contextual norms were not honoured is the Cambridge Analytica scandal. The contextual expectation of filling out a psychological questionnaire is certainly not that your data will be used for political advertisements. AI professionals should view the data they are using and the AI systems they are building through the lens of contextual norms.

6 Conclusion

Privacy considerations for AI professionals is extremely important as AI technology becomes more powerful and prevalent. There are serious consequences when privacy invasions occur. The Ethics Guidelines for Trustworthy AI (European-Commission, 2019) and the Privacy by Design (Cavoukian, 2009) are excellent resources AI professionals can leverage to ensure the privacy of users is upheld. While the current approach to tackle privacy is via privacy policies, instead AI professionals should view the topic through the lens of contextual integrity.

References

Ann Cavoukian. Privacy by Design - The 7 Foundational Principles. *International Association of Privacy Professionals*, 2009. (Online).

Nicholas Confessore. Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. *The New York Times*, 2018. (Online).

Jeffrey Dastin. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 2018. (Online).

European-Commission. Ethics Guidelines for Trustworthy AI. *Publications Office of the European Union*, 2019. (Online).

IBM. Cost of a Data Breach Report. *IBM: Computer Fraud & Security*, 2021. (Online).

Helen Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79:119–157, 2004.

Helen Nissenbaum. A Contextual Approach to Privacy Online. *Daedalus*, 140: 32–48, 2011.

Daniel J. Solove. A Taxonomy of Privacy. *University of Pennsylvania Law Review*, 154:477–564, 2006.

Daniel J. Solove. "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. *San Diego Law Review*, 44:745–772, 2007.

Samuel D. Warren and Louis D. Brandeis. The Right to Privacy. *Harvard Law Review*, 4:193–220, 1890.