Effect of Privacy Issues on User Behavior

Reasons and Impacts

Rohit Kumar Tilwani, Mysore Jagadish Vishwas Kaipu

Introduction

In the digital sphere, privacy refers to the right to have control and freedom for the use of one's own data. Privacy with respect to data is of great importance to users as it is linked to the values of (i) autonomy as in the control of what happens to one's information, how much information should be provided, and (ii) dignity as in people don't want everyone to know everything about them, which can make them vulnerable. Even with positive steps in this direction to conserve user control on their information, it is seen that people still have a sense of insecurity when it comes to providing information, even for many useful and genuine purposes. This also can be linked to anxiety, paranoia and an overall adverse impact on user behavior. We try to look upon the main reasons or actions which cause this insecurity and the reactions from users. This helps us to consider the problematic aspects while designing and developing any system that has a significant reliance on user data. These reasons can also be seen as a user feedback to understand what things are most important for a data-based system. Moreover, a byproduct of this analysis would be a development which will not only increase any user's trust in technology, but will also assist in taking a step towards more ethical systems which don't adversely impact user psychology.

Is there a reason to feel insecure !

There are more than enough reasons to make users unsure about sharing their data. A lot of it is attributed to the data leaks, breaches and incidents that have taken place. In the last 10 years, 300 data breaches that involve the theft of 100,000 or more records have occurred (Sobers 2021)[3]. Examples include Wonga Loans Breach (2017) (Anonymous 2017)[1], exposing bank details of more than 250,000 customers and Facebook data breach (2019) which affected 533 million users, and the personal details including phone numbers and emails were posted online by a hacking group (Holmes 2021)[2].

Along with these, secondary uses of data have also gathered a lot of negative attention. Use of data for anything else that it was gathered for constitutes secondary usage. An example to understand this would be taking user personal information on a healthcare application for the purpose of medical consultancy, and exploiting this data for analysis and to send adverts to users to take medical tests or supplements based on their condition, which is an example of surveillance capitalism. Facebook-Cambridge Analytica data scandal is well known for this where user data was used to affect election polls, without consent. These kinds of actions to optimize profits (whether monetary or not) at the expense of confidentiality severely undermine user privacy.

Unawareness is another factor that aggravates the already existing doubts in users (Guerrero 2019)[7]. While collecting relevant data for the systems, it is our responsibility to clearly identify and communicate the purposes of doing so. There have been many instances where mobile applications have been found to steal user data unethically, including personal phone data and bank details (Seckhose 2021)[4]. However, taking user consent on incomprehensible and twisted privacy policies, even if legally correct, don't help a layman understand their data usage and hence don't develop confidence and trust between data providers (end users) and data collectors (applications, systems, and organizations) (Rao and Dwivedi 2017)[6].

Impacts on user behavior

For the above actions, there are reactions which are not good for the general public and also for data collection for useful purposes. People avoid giving useful data or give wrong data which can hinder research and falsely project the situation. This can be linked to self-censorship. For instance, people not filling out the online psychiatrist form with all information, fearing some of it might end up in the wrong hands, and in this process might also miss something which is critical to share. Fear of data breaches goes to the extent that many people are reluctant to use a bank card having large amounts of money, fearing cyber-theft, or don't trust IoT devices in their houses (Krauth 2017)[8]. These days, consumers are revealing less and less personal information, this can also be understood by the nature to unsubscribe newsletters, making the profiles private over social networking platforms (Guerrero 2019)[7], and not allowing permissions to mobile applications.

According to a survey mentioned in Swant 2019[5], the percentage of people willing to share their home address went down from 41% to 31% from 2018 to 2019, 41% to 33% in relation to sharing the name of the spouse, and 54% were willing to share their email address, as compared to 61% last year. Usage of data for purposes not clearly stated or not clearly understood by the users makes them feel as if they are being seen all the time, for e.g. if a user is not communicated clearly that their social media advertisements are influenced by what they search for, they will be surprised and in a negative way, on noticing such adverts on their social media which might also be related to something they don't want to share, and this might alter a user's natural surfing behavior, under the impression that they are constantly being watched and can be identified. Another example is that in 2005, when Facebook began, more than 80% of its users disclosed their private information. With the passage of time, disclosure had decreased to less than 20% by 2011 (Acquisti et al., 2015).

Conclusion

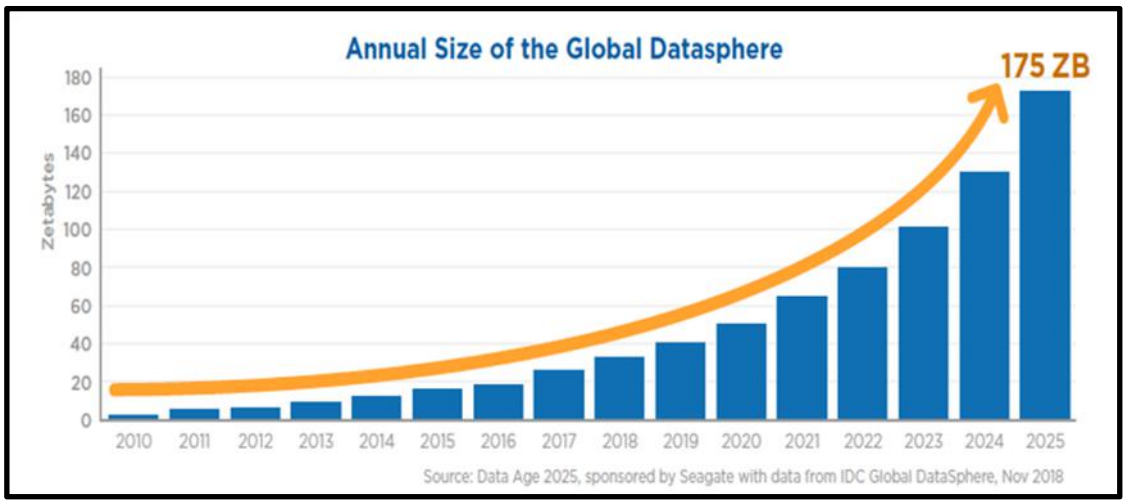
The scenario of the current digital world is having a significant impact on users, who as a result are getting more and more reluctant to share data and decreasing their trust on modern systems. In order to gain user confidence, data collection and usage policies should be more transparent and more user inclusive to increase awareness. Users should be allowed to easily opt-out of data sharing and be notified if their data is

breached, and should be helped to deal with it, rather than denying breaches. Purposes should be clearly communicated in an understandable manner so it doesn't come off as a surprise.

Part-2: Understanding of the ethical issue in academic literature

Problem of Privacy in the world of Artificial Intelligence

Market for Artificial Intelligence applications is increasing rapidly due to its applicability in a variety of domains. As a greater number of applications are being developed, the need for generating and collecting data is also increasing. This can be evident by seeing the below plot of a recent survey projecting the data that the AI applications are generating every year.



This big data driven paradigm has dominated artificial intelligence research in recent years, resulting in a new wave of AI advancement (Li and Zhang, 2017)[10] .Machine learning algorithms are primarily used to perform big data analytics and to mimic human cognitive capabilities. There are mainly three types of machine learning algorithms in AI: 1) Supervised, 2) Unsupervised, and 3) Reinforcement learning algorithms.(Bostrom and Yudkowsky)[13]

Most of the industry applications currently are based on Supervised Machine learning algorithms, these algorithms rely heavily on training data sets, the performance of the machine learning algorithms is provisory on the size and quality of the dataset which is being provided, so most of the AI applications are exceedingly dependent on big data.(Dignum 2018)[12]

Once such type of application which uses the capability of AI is Facebook, Facebook access users' personal information as a training dataset to their underlying machine learning algorithm, this information which is shared constantly from user to the company had caused a dilemma about of how millions of people's private information was being used in the application. The data of a user is revealed through the user's actions on the internet —both knowingly and unwittingly—to one another, to commercial entities, and to our governments (Acquisti, A., Brandimarte, L. and Lowestein, G. 2015)[9]. The data collecting focuses on continuous improvements in the ability to

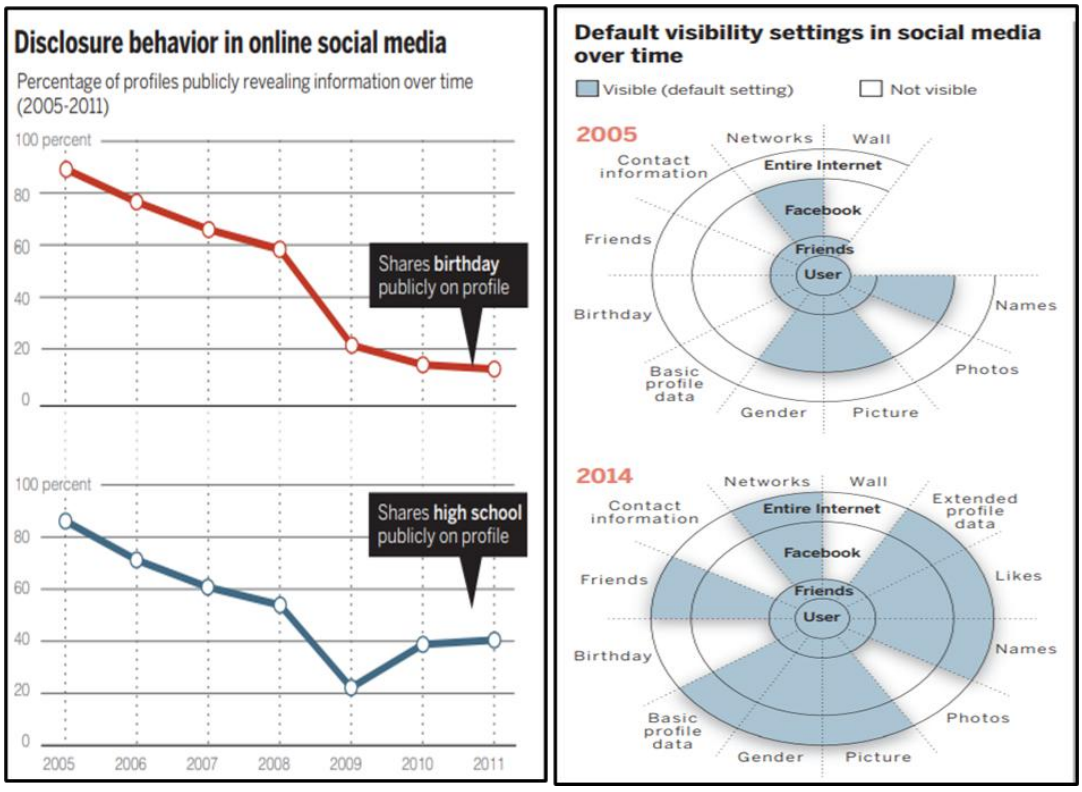
combine, analyze, and make sensitive judgments from people's data. People establish their private and public information in various ways, comprising the desire for protection from societal influence and control, as well as the need for companionship and psychological rest (Acquisti, A., Brandimarte, L. and Lowestein, G. 2015)[9]. Please watch the below one-minute video of dialogues exchanged between the Facebook CEO and an American senator when questioned about his privacy / sharing private information.

https://www.youtube.com/watch?v=qJeySsOR4Bk&ab_channel=SouthChinaMorningPost

The above video is an example of how information privacy can make the general public conscious when questions about private life are asked directly but the people are not so conscious when the information is shared on their social media pages.

This example is evidence showing how society views an artificial intelligence application. The companies try to make their product better by acquiring more information from their users. This practice of acquiring more information to make an application / product has led privacy and trust of users to grow distant from each other, whereas they should be going hand-in-hand.

On the other hand, an artificial intelligence application requires large and quality datasets, so for a company to provide a better service to the users they need to collect the user's information and feed it to the algorithm.



Source: Acquisti et al., 2015

The above graph shows the behavior of the society / user's on sharing their information overtime.

The study of actual disclosure behavior of online social network users highlighted that over time, many users increased the amount of personal information revealed to their friends (those connected to them on the network) while simultaneously decreasing the amounts revealed to strangers (Acquisti, A., Brandimarte, L. and Lowestein, G. 2015)[9].

Over the time Facebook included many other fields / different data types to increase their market and product share in the industry that led to massive change in the user behavior (Acquisti, A., Brandimarte, L. and Lowestein, G. 2015)[9].

The problems arising in behaviors of users with respect to their privacy which arise in AI applications are:

1) Dilemma of Data

The main dilemma for users and AI practitioners is; how much data is required by these smart systems which is enough to make an accurate prediction? This is the question which can address the confusion between privacy and AI application development.

2) Privacy in Data amassing

The data collected from your smart devices such as chatbots, smartphones, laptops of you and your family can be stored for longer periods (even decades) on the databases.(Li and Zhang, 2017)[10] If this data which is stored is used properly, it can make lives of you, your family and even the entire society better but it is not guaranteed to be used so, the information can be used for commercial purposes illegally by the technological companies. The data produced by your smartphones are from the electrical activities, such as credit/debit notes, geographical coordinates and travel routes, which involve personal sensitive information, if this information is used illegally, it can harm the lives of many people in an instance.On the other hand, Information collected about a person's health and stored for a long term can help the application make better predictions and help save people's life.So, the data acquisition should be limited by requirements of an application else may cause privacy invasion.(Li and Zhang, 2017)[10]

3) Privacy in on-demand computing

Lot of companies (including big and small companies) are making use of cloud computing capabilities, they are migrating their data to the cloud as it's a low-cost, simple-to-use, and convenient way to receive on-demand network access to a shared pool (Li and Zhang, 2017)[10]. When the user's data is migrated from local databases to a cloud where it can be used by many other applications/companies, the user's privacy needs to be ensured. Utility computing has been configured as the major infrastructure of many AI applications due to the increasing compute requirements of artificial intelligence, therefore privacy issues are something we should consider while employing such intelligence techniques (Li and Zhang, 2017)[10]

4) Privacy in Knowledge Extraction

As the world of AI applications are increasing, the tools for extracting user's information have also exponentially increased.(Li and Zhang, 2017)[10] These tools are used in variety of applications and websites, they collect the data of a user in fragments. These

fragments can later be combined to analyze the pattern in the user's behavior. Based on the pattern of the user, the technology companies promote their ads and treat it as commercial data instead of personal data.

In contrary, the knowledge extraction tools present on the websites keep traces of your purchase history, websites you have visited and other records, then all these records are integrated together When internet site visit traces, purchase processes, and other sorts of record data are integrated, for example, you can create a person's behavior map and assess their specific preferences and behavioral tendencies, allowing AI application to better predict consumer demands. As a result, businesses are able to supply them with the essential information, possible products or services, ahead of time. Personalization has emerged as a key element and highlight of today's sophisticated applications. However, these tailored modification procedures come with the exploration and exploitation of a user's personal and sensitive information.(Li and Zhang, 2017)[10]

So, we can say that the main concern Privacy in knowledge extraction is how to avoid privacy invasion by these knowledge extraction tools.

Possible elucidation for privacy problem

The articles "**PRIVACY AS CONTEXTUAL INTEGRITY**" by Helen Nissenbaum and **"An Exploration on Artificial Intelligence Application: From Security, Privacy and Ethic Perspective"** by Xiuquan L and Tao Zhang outlines the practices that address the issue of insecurity in data privacy.

The article by Helen Nissenbaum proposes "contextual Integrity" as an alternative benchmark for privacy to properly highlight the essence of the information technology difficulties. Contextual integrity relates proper privacy protection to specific context norms, requiring that data collection and dissemination be acceptable for that context and follow the prevailing distribution rules within it.

It also outlines three principles for contextual integrity:

1) **Principle 1:** *Protecting Privacy of Individuals Against Intrusive Government Agents* (Nissenbaum 2004)[14]

2) **Principle 2:** *Restricting Access to Intimate, Sensitive, or Confidential Information* (Nissenbaum 2004)[14]

3) **Principle 3:** *Curtailing Intrusions into Spaces or Spheres Deemed Private or Personal* (Nissenbaum 2004)[14]

Principle 1 addresses the dilemma of data amassing by government agents and big technology companies who collect and use personal information in the name of "National Security". This principle can be viewed as a subset of the more powerful, more universal idea of defending individuals against unjust government dominance. Individual privacy is thereby preserved through the application of broad, well-defined, and widely recognized political norms addressing the balance of power, which, among other things, limit government intrusion into people's lives and liberties (Nissenbaum 2004)[14]

Principle 2 addresses the issues of privacy in knowledge extraction where the knowledge extraction tools are restricted from collecting intimate, sensitive and confidential information fragments of the individual. This principle restricts the collection of fragmental information, if the information fragments is intimate, sensitive and confidential then that knowledge should not be extracted or there should be no pattern that should be created by using this information which can invade the privacy of a user.(Dignum 2018)[12] It does not concern about who is collecting the information, the principle focuses more on the content of the information which is collected. This principle helps us to differentiate between sensitive information and other general information.(Nissenbaum 2004)[14]

Principle 3 address the issue of cloud computing, where the user's data is treated as the user's home / private space, which is the user is sovereign in his/ her own domain; here user is the place where the user's information is stored, when people are inside their own private spaces, this concept appears to promote a presumption in favor of protecting themselves from the eyes of others. This principle allows the user to contest and have rights for his/her information which are stored in the cloud and can have control on the way the data is shared or used by any third party. He has all the right to share or not share his information which is on cloud even though the cloud space may not belong to him.(Nissenbaum 2004)[14]

The article by Xiuquan L and Tao Zhang helps us understand the countermeasures that can be used to solve security, privacy and ethical aspects in the world of Artificial intelligent applications.They emphasis on three countermeasures for solving the privacy issues, they are:

1) Emphasizing Safety, Privacy and Ethic Research(Li and Zhang, 2017)[10]
2) Strengthening Regulation on Al Development(Li and Zhang, 2017)[10]
3) Improving Security and Privacy by Utilizing Al Technology(Li and Zhang, 2017)[10]

The first countermeasure of emphasizing on safety, privacy and ethical research focuses on embedding ethical rules while designing an AI application, making the AI application more transparent and explainable to the general audience and focuses on improving the security that is provided to the data of the user used by the AI systems.(Li and Zhang, 2017)[10]

The second countermeasure for regulations on AI development focuses more for the lawmakers, it provides insights on how the Law and policy making should keep up with the rapidly growing AI systems, it discusses about the standard rules which have to be kept in mind while designing or developing any AI application and once the AI application is developed how the application must be managed and supervised by its creators.(Li and Zhang, 2017)[10]

The third countermeasure gives us the methods to protect the privacy and enhance the security of the systems. AI has the potential to greatly improve not just our society's and cyberspace's security, but also individual privacy protection. The smart systems can be

used for securing a person's life (like facial recognition system to enter a property and so on) , enhancing cybersecurity to protect the users information present in the cloud and improving privacy protection by desensitizing the collected data and creating a threshold for data disclosure , leading to healthy evolution of AI applications in the society.(Li and Zhang, 2017)[10]

Part 3: Process of Resource Creation

In this section, we describe the process followed to create this entire resource. The purpose has been to convey more relevant information and also maintain the readability i.e. without making the resource too long to follow. We cover who our intended audience is considering their background, and what can be the preferred way to deliver and communicate this information effectively, in order to have intended effect. Furthermore, what were the reasons to choose this topic that focuses on users, why this format was chosen and why we should talk more on this and similar issues also.

Firstly, the target audience for our topic of discussion are the system designers, architects and developers or basically the creators who design and develop data based systems or products for users. This is because although educating users is important which will make them aware of issues related to privacy and data, it is the creators who need to understand such aspects during ideation or implementation phases in order to create an ethical system embedded with values that doesn't negatively impact consumer behavior.

This topic can be a part of the initial training for the targeted audience, i.e. who are supposed to have a contribution in systems development. This can also be communicated in the form of compliance training at organizations, at meetings and conferences, and be made available online in the form of blogs etc, to cover as much relevant audience as possible. The reason being that the importance of consumers being prime stakeholders and their well-being an exigent part during the entire process should be ingrained in the core principles of any model and method.

The resource presents and communicates to architects and developers an analysis of issues that need to be solved in the context of data usage and data security, the extent of transparency a design should have, involving consumers into the process and helping them make the right decisions. The importance of data collection and usage while developing systems is well understood by the targeted audience, along with how user data can be harnessed for several purposes, for e.g. to improve user experience or to send advertisements. With the audience having familiarity with system functioning and methods, the examples convey the form in which ethical issues currently prevail in our systems and what are the main factors for that. It also helps them see things from the point of view of users and as a feedback that can help in further improvement of the processes followed.

It is evident from the overall materials available online that privacy issues, data breaches, and software flaws and faults are much talked about along with how to perform damage control and disaster management. In all these talks of technology systems and estimations of value of compromised data, it is often forgotten how the victims are affected in terms of much more than just the loss of their data, which in many cases can be more than financial loss in numbers. For instance, someone can be blackmailed or have their personal information leaked or posted online, such instances can deeply impact anyone's mental and emotional health. With this topic, we try to make an attempt to reflect the user side of the story and how the issues are perceived from the other end. As mentioned above, this is something which is not very often talked about and hence it is challenging to find a lot of relevant academic literature in this respect. However, a number of headlines and verified news based on personal experiences of victims and surveys can be obtained online. Along with academic literature, we have tried our best to incorporate the knowledge from all such verified sources. The resource is a written format, which is more of a brief document yet properly explains the situation and provides condensed material for quick reference, making it easy for everyone to read or treat as a handbook and get a good idea of existing issues, rather than going through entire research papers. The format employs a method of providing a suitable example to convey the point and also provide actual occurrences of such instances, helping the reader to relate to.

After creating the resource, we also tried to identify its limitations as a step for potential improvements, which if possible can also be included in the document. Something that we identified is more of a trade-off of detailing vs compactness, rather than an improvement, which is that although the document is focused upon describing one issue (impact on user behavior) in some depth, this can be extended in two ways, one of which is to cover more issues (more breadth) in same amount of material but will result in the resource being just a introduction of various issues and won't be suitable for the intended audience, who already have some understanding of issues. Another way would be to cover more issues in depth, which although would be very helpful but will impact the readability and compactness of the overall document, making it hard to follow and defeating our purpose of making a resource which is dense in information, i.e. covering more in fewer words and also easier to communicate.

References

[1] Anonymous (2017). Wonga data breach 'affects 245,000 UK customers', BBC News, Available at: https://www.bbc.com/news/business-39544762

[2] Holmes, Aaron (2021). 533 million Facebook users' phone numbers and personal data have been leaked online, Insider, Available at: htttps://www.businessinsider.com/stolen-data-of-533-million-facebook-users-leaked-online-2021-4?r=US&IR=T

[3] Sobers, Rob (2021). 98 Must-Know Data Breach Statistics for 2021, Varonis, Available at: https://www.varonis.com/blog/data-breach-statistics

[4] Seckhose, Marcia (2021). Android apps that steal banking details were downloaded 300,000 times in just 4 months, Business Insider, Available at: https://www.businessinsider.in/tech/apps/news/android-apps-that-steal-banking-information-were-downloaded-300000-times/articleshow/88022001.cms

[5] Swant, Marty (2019). People Are Becoming More Reluctant To Share Personal Data, Survey Reveals, Forbes, Available at: https://www.forbes.com/sites/martyswant/2019/08/15/people-are-becoming-more-reluctant-to-share-personal-data-survey-reveals/?sh=1f64613c1ed1

[6] Rao, Vikram and Dwivedi, Kruttika (2017). To share or not to share, Deloitte, Available at: https://www2.deloitte.com/us/en/insights/industry/retail-distribution/sharing-personal-information-consumer-privacy-concerns.html

[7] Guerrero, Daniel (2019). Personal data consent: What customers are afraid of and what do they look for?, Expilab, Available at: https://expilab.com/personal-data-consent-what-customers-are-afraid-of-and-what-do-they-look-for/

[8] Krauth, Olivia (2017). Only 9% of consumers fully trust IoT devices, but many refuse to disconnect, TechRepublic, Available at: https://www.techrepublic.com/article/only-9-of-consumers-fully-trust-iot-devices-but-many-refuse-to-disconnect/

[9] Acquisti, A., Brandimarte, L. and Lowestein, G. (2015) Privacy and human behavior in the age of information. Science, vol. 347(6221), pp. 509-514, Available at: https://www.cmu.edu/dietrich/sds/docs/loewenstein/PrivacyHumanBeh.pdf

[10] Li, Xiuquan and Zhang, Tao.(2017). An Exploration on Artificial Intelligence Application: From Security, Privacy and Ethic Perspective, 2nd IEEE International Conference on Cloud Computing and Big Data Analysis

[11] Stahl , Bernd, Carsten, and Wright, David,. (2018) Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation , IEEE computer and Reliablity societies

[12] Dignum , Virgnia.(2018) Ethics in artificial intelligence: introduction to the special issue, Springer Science+Business Media B.V.

[13] Bostrom, Nick and Yudkowsky,Eliezer . The Ethics of Artificial Intelligence ,In Cambridge Handbook of Artificial Intelligence, edited by Keith Frankish and William Ramsey. New York: Cambridge University Press.

[14] Nissenbaum,Helen,. (2004) Privacy as contextual integrity ,Washington Law Review Association