ISSN (e): 2250 – 3005 || Volume, 12 || Issue, 2|| Mar. - Apr. – 2022 ||
International Journal of Computational Engineering Research (IJCER)

AI and data privacy in ConTech: A case study of viAct's responsible scenario based AI

Gary Ng[1], Hugo Cheuk[2], Surendra Singh[3], Barnali Sharma[4], Baby Sharma[5]*
118 Wai Yip St, Kwun Tong, Hong Kong
[4] Corresponding Author: Baby Sharma (baby.sharma@viact.ai)

Abstract
Most of the AI-based applications are built on the foundation of deep learning methods. In the ConTech ecosystem, large scale user data are collected through vision intelligence which is proportionally used for training the AI for various scenarios. In this context, it is observed that for building any AI module, a massive data collection is the prime necessity for deep learning which accompanies inevitable privacy issues. Highly sensitive user data such as photos and videos are indefinitely with the companies which collect them and user cannot delete it or restrict its usability. Thus, vision intelligence powered AI which is popularly used in the ConTech ecosystem are potential subject to legal and privacy matters. GDPR regulations are stringent in this sector because with the inclusion of AI in the construction sector there is a rise in risks of privacy damages. However many startups and large companies have set good example of accuracy maintaining privacy norms together. viAct (Hong Kong) is one of such world class scenario based ConTech startups known for its privacy ensuring platform. viAct has taken steps like blurring and masking of human faces, encryption of stored data, privacy preserving deep learning for computer vision and edge AI for computing in order to mitigate such privacy issues. The presented case study of viAct's AI thus showcases a good example of responsible AI.

Date of Submission: 18-04-2022 Date of acceptance: 03-05-2022

I. Introduction

In basic terms, privacy is the right not to be observed. In day to day basis, many subtle activities such as wearing sun glasses, shutting doors that are done by humans in order to moderate their privacy. With technology intervention, people have become more concerned about being in a manner so as to enhance their own privacy. Privacy is of extreme necessity due to many reasons: it allows people to better calculate their behavior, to make their non-coerced decisions, be strategic in their social interactions, and also to take decisions and actions that do not conform to certain social norms. Every time one searches something on internet, browses any websites, or when one use mobile apps, the data is either given out explicitly or without knowledge. In most of the times, one gives away rights to collect and process one's data legally by clicking on "I agree" button of terms and conditions of using services. Much information like Name, Age, Emails, Contacts, Videos, or Photo uploads is explicitly submitted while other information like browsing behavior, clicks, likes and dislikes are analyzed for understanding data to improve customer services. Such data are often not for sale but sometimes, there have been instances where third party companies have scrapped sensitive user data through data breaches. In the growing internet world, there have been instances where sole intention of many companies is to collect user data. This is often done by luring consumers with their online services, later the collected data is sold to third parties against vast amounts of money.

The situation of privacy infringement has worsened in the recent times, especially after the surge of Artificial Intelligence and Machine Learning in the digital world. Many malicious mobile applications are often made for the purpose of collecting data without seeking any permission from the user. These are generally disguised as gaming or entertainment apps and are major sources of data privacy infringement. In today's world, it is often important to protect very sensitive data like videos, personal images, call history, GPS location, messages, from getting stolen by malware apps. Data privacy-awareness has been a buzz word these days everywhere. Data privacy or information privacy, in simple words, is concerned with the proper handling, processing, storing, as well as using the personal data/information of the individuals, companies, etc. Data

security focuses on protecting the personal data from any unauthorized third-party access, malicious attacks and/or exploitation of data.

Every industry is vulnerable to privacy and cyber security risks, the construction industry being no exception to it. In fact, the construction industry is a ripe target for attacks owing to its lucrative nature – the $10 trillion sector, which is one of the largest in the world – coupled with increasing vulnerability. Thus the current paper presents a case study of Asia's first scenario based responsible AI and how it sticks to data privacy norms to fulfill GDPR compliances.

II. Review of literature

2.1 AI & Data Privacy

AI is one of those businesses that need lots of data from users/consumers. The accuracy of AI depends on training of the AI with appropriate data sets. This in turn calls for requirement of enormous data collection which may sometime hinder privacy of users. This has led AI to be more and more focused and accurate about customers (McCarthy, 2017). On the other hand, these businesses are being strictly regulated by governments to provide certain limits as they are becoming more and more invasive into public's privacy. Thus, AI bases business often try to collect and use public data and information to find out information while remaining in compliance with regulatory rules. Last few years have seen numerous good governance guidelines on which the trustworthy AI was published. Most of these AI governance frameworks mostly agree to the following: privacy and data governance, accountability and auditability, robustness and security, transparency and explainability, fairness and non-discrimination, human oversight, and promotion of human values (Wadlow, 2018). Some prominent examples of responsible AI frameworks by public organizations include the UNESCO's Recommendation on the Ethics of AI, China's ethical guidelines for the use of AI, the Council of Europe's Report "Towards Regulation of AI Systems", the OECD AI Principles, and the Ethics Guidelines for trustworthy AI by the High-Level Expert Group on AI set up by the European Commission. Apart from this, many self-regulatory initiatives by companies have been put forward to build responsible use of AI. Therefore, it seems that the question of whether Artificial Intelligence systems will be legally liable depends on at least three factors: the limitations of AI systems and whether they are communicated to the purchaser; whether an AI system is a product or a service; whether the offence requires a mental intent or is a strict liability offence. If an Artificial Intelligence system is held liable, the question arises of whether it should be held liable as an agent, an accomplice, or a perpetrator (Debney, 2018). The EGE (European Group on Ethics in Science and New Technologies) has proposed a set of basic principles and democratic prerequisites for AI to be used in construction. This is based on the fundamental values laid down in the European Union Treaties and in the European Union Charter of Fundamental Rights (Clavero, 2018). This is the first step that has been taken in this sector for the formulation of ethical guidelines for maintaining a golden standard for use of AI in construction ecosystem. It has been stated that "human dignity which is represented as the recognition of the inherent human state of being worthy of respect" must not be violated by 'autonomous' technologies. It also emphasizes on the standardizing some (legal) limits the way in which people perceive while dealing with automatic devices that they are not humans rather are just smart machine and algorithms. The relational conception of human dignity which indicates that our social relations asks to ensure that one is aware that whether and when we are interacting with a machine or another human being, and that we reserve the right to vest certain tasks to the human or the machine (Bartneck et al., 2021). The second is "Autonomy". The principle of autonomy implies the freedom of the human being translating into human responsibility and thus bringing control over and knowledge about 'autonomous' systems and the fact that "they must not impair freedom of human beings to set their own standards and norms and be able to live according to them" (Debney, 2018). Third point of consideration is "Responsibility which means 'autonomous' systems must only be developed and used for purpose which fulfils some or the other social and environmental good as a result of a deliberative democratic processes (Booth et al., 2018). Moreover, there has been a raising concern about preventing, reporting and neutralizing the discriminatory biases in data sets which are used in training and running the AI systems.

2.2 Privacy issues in using AI in construction

Recent times have seen an upsurge in the field of ConTech ecosystem. This has simultaneously raised new concerns such as an appropriate allocation of risk, privacy damage, cyber risk, risk of overreliance on technology, legal liability of the robots, the difficulty of migrating to a new supplier, tort and breach of contract and warranty,

Moreover the major issue is privacy of data and its misuse. There have been may privacy issues reported in the past regarding privacy in construction. A study conducted by IBM revealed that 74% of the construction-related organizations are not prepared for cyber-attacks and do not have an incident response plan in place. The study conducted by Safety Detectives revealed that the construction industry was the third most common industry to have experience ransomware attacks in the year 2021. The 2020 Forster survey revealed

that 75% of the respondents in the construction, engineering and infrastructure industries have experienced cyber-incident in the year 2019.

All companies are vulnerable to cyber security risks, but it is more so in case of the construction industry due to the following reasons: The industry is highly unregulated when it comes to privacy and cyber security. The transactions in the industry contain significant amount of personal information as well as sensitive business data, particularly financial data, which attracts the attackers. The construction companies work with a large number of vendors, and as such each transaction may involve multiple parties, thus providing ample opportunities for the attackers to wreak havoc. In the recent years; the industry has been adopting new technologies like AI, robotics, etc. Given their interconnectivity, greater data security and privacy considerations are thus required.

2.3 GDPR Guidelines and data privacy through AI

Looking into the growing risks of cyber-attacks and increasing privacy concerns, the governments all over the world have been coming up with new and stricter data protection laws, that dictate the manner in which companies handle data and incorporate values to strive in the market. One such landmark data privacy law is the General Data Protection Regulation or the GDPR. The General Data Protection Regulation (GDPR) is new data privacy and security law drafted and passed by the European Union (EU) that imposes a set of obligations onto organizations across the world, so far as they collect data related to people in the EU. The GDPR is the toughest privacy and security law in the world and was put into effect on May 25, 2018. Any organization violating the GDPR privacy and security standards shall be levied huge fines that may range to tens of millions of Euros.

At a time when more and more people are entrusting their personal data with cloud services, and breaches have become an everyday happening, introduction of GDPR reflects Europe's firm stance on data privacy and security. The regulation itself is of very high standard to meet, requiring the companies to invest large sums of money to ensure they are in compliance. The entire text of GDPR consists of 99 articles, setting out the rights of individuals and obligations imposed on the businesses that are subjected to regulations. Under GDPR, any company that stores or processes personal information about EU citizens within EU states must comply with the GDPR, even if they do not have a business presence within the EU.

GDPR addresses and covers the following types of data:

- Personally identifiable information, including names, addresses, date of birth, social security numbers
- Web-based data, including user location, IP address, cookies, and RFID (Radio-frequency Identification) tags
- Health (HIPAA) and genetic data
- Biometric data
- Racial and/or ethnic data
- Political opinions
- Sexual orientation

Further, GDPR contains the following rules and regulations regarding:

Data Protection: If any company processes data of any EU citizen(s), it must do so in accordance to the seven protections and accountability principles laid down by GDPR, namely: lawfulness, fairness and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; and accountability.

Accountability: According to GDPR, the data controllers should be able to demonstrate that they are GDPR compliant. They can do this by:

- Designating data protection responsibilities to their team;
- Maintaining a detailed report of the data that has been collected, how the data has been used, where has it been stored, which employee(s) is(are) responsible for it, and the like;
- Train the staff and implement technical and organizational security measures;
- Have Data Processing Agreement contracts with third parties which processes data for the concerned company(ies)
- Appoint a Data Protection Officer (though not all organizations require one)

Data Security*:* Companies are required to handle data securely by implementing "appropriate technical and organizational measures". Technical measures may include anything from using two-factor authentication on accounts where personal data has been stored to end-to-end encryption of cloud platform. Similarly, organizational measures may include things like staff training, data privacy policy, limiting access to personal data to only those employees that require it, and the like.

Apart from the above, if any company faces a breach of data, it has to tell the data subject(s) within 72 hours or face penalties.

Thus, the increasing public concern over privacy on the business sector has resulted in stringent rules like the GDPR that regulate the way in which the companies use the personal data of the individuals. Now, companies in order to function in these countries need to abide by these regulations or else be subjected to heavy fines.

III. Case study: viAct a good example of responsible AI in construction

viAct is a ConTech startup from Hong Kong that provides "Scenario-based Vision Intelligence" solutions exclusively for construction industry all across Asia & Europe by successfully deploying around 50 sites. viAct's smart AI modules has been successfully providing extremely granular insights on environmental compliances in construction jobsites by not only tracking objects but by transforming vision to practical actions. In this notion, viAct's scenario-based AI is playing a significant role in measuring, monitoring, tracking, predicting and reducing carbon emissions. It has also been constantly striving to help the construction industry in tracking carbon credit emission and carbon credit monitoring and auditing. In the present times, with the continuous increase in carbon emissions, it has become pertinent for the construction companies to reduce their carbon footprint and erect buildings that are not just economical but also environmentally sustainable.

At viAct, it is to maintain privacy of data, not only to keep the company safe from penalties and fines, but also as a part of our moral responsibility. viAct believe that privacy of data cannot be compromised at any cost. This is why viAct has presented a great example of responsible AI by being very vigilant when it comes to protection and privacy of data of clients. viAct ensures data privacy of its clients in the following ways:

3.1 Blurring or masking human faces

It is important to have privacy-aware image recognition in construction sites as faces are ubiquitous in datasets. Even those people, who are not targets directly, come under the camera surveillance being public datasets. Thus, to abide to GDPR compliances faces are blurred

It is done in two phases. In the first phase, prominent type of private information such as faces (or even number plates of dump truck/ construction vehicles) are annotated by automatic face detectors (such as Amazon Rekognition) to detect the faces. Later accuracy of face detection annotations is obtained using crowd sourcing on Amazon Mechanical Turk which helps in refining of results. In subsequent step, blurring is used as a method for privacy preservation by obfuscating sensitive image areas. This leads to the creation of face-blurred version of the obtained dataset. A minimal impact on accuracy is observed using the face obfuscation models. It has been observed that the drop in accuracy is as low as Benchmarking with multiple deep neural networks on face-blurred images showed that the overall recognition accuracy drops only slightly (≤0.68%).

Fig 1: Original picture from construction jobsite

Fig 2: Face blurring by obfuscating sensitive image areas like face for construction workers

3.2 Encryption of the stored data

viAct stores collected data in high security cloud computing platforms like Amazon Web Services for ensuring greater safety of clients' data. The Amazon Machine Image (AMI) which are backed by Amazon EBS snapshots take advantages of Amazon EBS encryption. Encryption and attachment to an AMI are done both at snapshots of both and root volumes. The following diagrams depict examples of instances launching from AMIs using non-default encryption parameters.

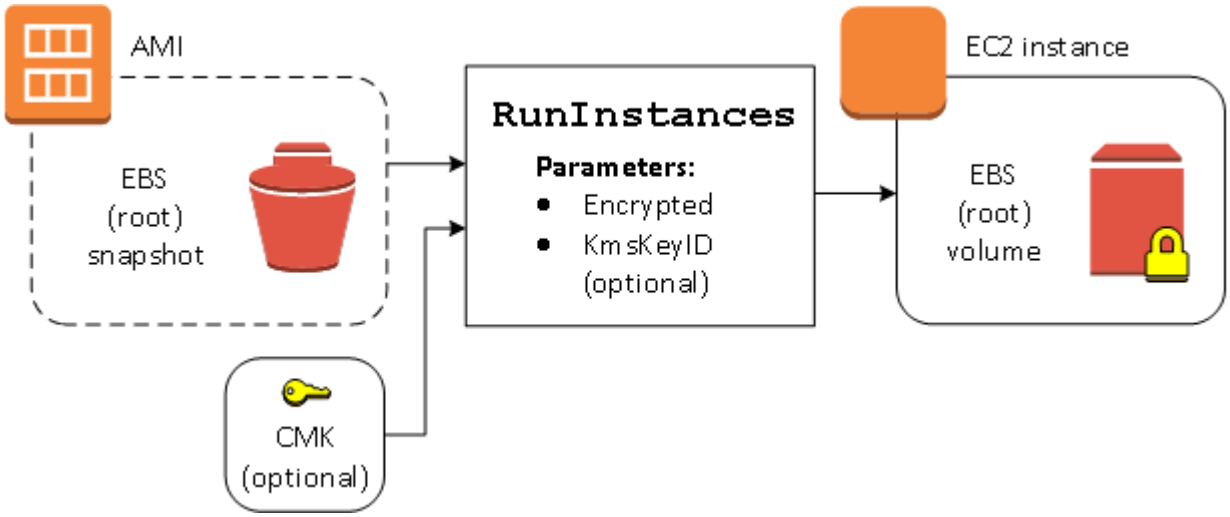

Fig shows: *Encrypting a volume during launch*- An AMI backed by an unencrypted snapshot is used to launch an EC2 instance with an encrypted EBS volume.

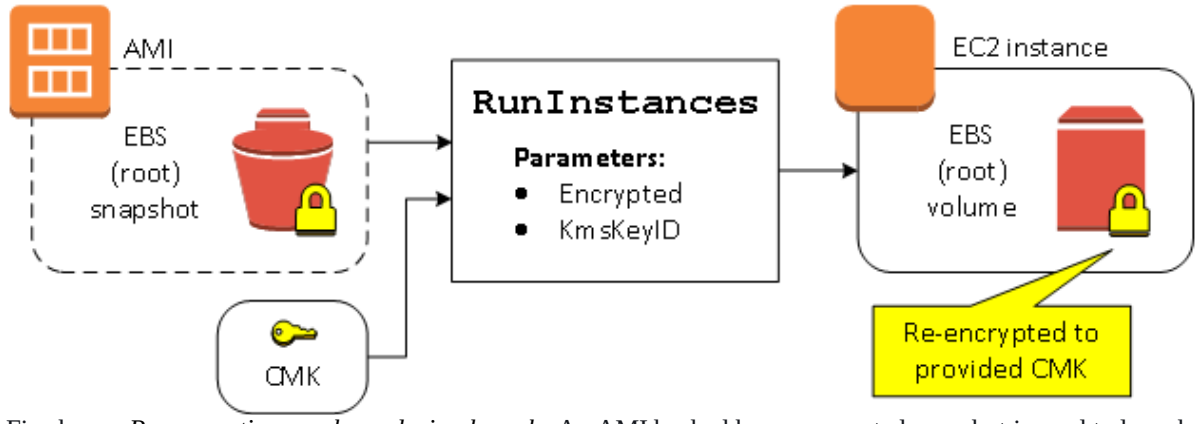

Fig shows: *Re-encrypting a volume during launch*- An AMI backed by an encrypted snapshot is used to launch an EC2 instance with an EBS volume encrypted by a new KMS key.

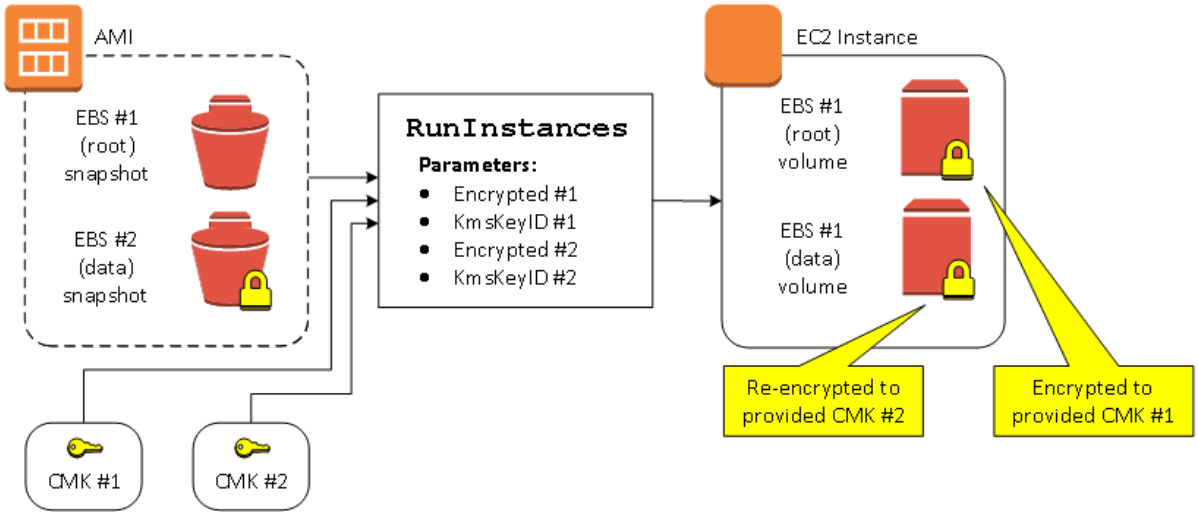

Fig shows: *Change encryption state of multiple volumes during launch*- In this more complex example, an AMI backed by multiple snapshots (each with its own encryption state) is used to launch an EC2 instance with a newly encrypted volume and a re-encrypted volume.

3.3 Privacy-preserving Deep Learning for Computer Vision

The dataset used in machine learning is generally private, which is not accessible to everyone and can only be accessed by dataset owner. Hence even for the trained modules, privacy-preserving machine learning is concerned with adversaries which are trying to infer private data. The most general approach which is commonly used to defend against such attacks such as Model inversion attacks, Membership inference attacks and Training data extraction attacks is Differential Privacy (DP). DP offers strong mathematical guarantees of the visual privacy of the individuals whose data is contained in a database.

3.4 Edge AI for computer vision

Edge AI allows processing sensitive data locally, without the need to send all video streams into the cloud (data-offloading) for processing it there. viAct operates computer vision with edge computing in order to run highly efficient and private on-device machine learning (Edge AI). Due to the decentralized data processing near the data source, numerous limitations of data privacy in image processing can be overcome by Edge AI.
The input video data is processed locally in the connected devices by Edge Computing. The anonymous metadata is sent to the Cloud after all the visual data provided by a camera is analyzed at the edge.

IV. Conclusion

Most of the AI-based applications are built on the foundation of deep learning methods. In the ConTech ecosystem, large scale user data are collected through vision intelligence which is proportionally used for training the AI for various scenarios. In this context, it is observed that for building any AI module, a massive data collection is prime necessity for deep learning which accompanies inevitable privacy issues. Highly sensitive user data such as photos and videos are indefinitely with the companies which collect them and

user has cannot delete it or restrict its usability. Thus, vision intelligence powered AI which is popularly used in the ConTech ecosystem are potential subject to legal and privacy matters. GDPR regulations are stringent in this sector because with the inclusion of AI in the construction sector there is a rise in risks such as privacy damage, cyber risk, risk of overreliance on technology etc. However many startups and large companies have set good example of accuracy and privacy together and viAct (Hong Kong) is one of those. viAct's world class scenario based AI has been known for its privacy ensuring platform. viAct has taken steps like blurring and masking of human faces, encryption of stored data, privacy preserving deep learning for computer vision and edge AI for computing to mitigate privacy issue. The presented case study thus showcases a good example of responsible AI.

References

[1]. Wadlow, T. 2018, Feature How Artificial Intelligence Supports Construction Industry, Construction Global weekly, https://www.constructionglobal.com/equipment-andit/feature-how-artificial-intelligence-supports-construction-industry.
[2]. McCarthy, J. 2017. What is artificial intelligence?, Stanford University. http://wwwformal.stanford.edu/jmc/whatisai/
[3]. Disruptive technologies: Advances that will transform life, business, and the global economy, McKinsey Global Institute, May 2013.
[4]. Booth, K. 2018, The Impact of Artificial Intelligence in the Construction Industry http://www.bdcmagazine.com/the-impact-of-artificial-intelligence-in-the-constructionindustry/.
[5]. Wadlow, T 2018, Feature: How Artificial Intelligence Supports Construction Industry, Construction Global weekly, https://www.constructionglobal.com/equipment-andit/feature-how-artificial-intelligence-supports-construction-industry.
[6]. Debney, P 2018 How artificial intelligence is changing the construction industry, Artificial intelligence news, https://www.artificialintelligence-news.com/2018/03/16/howartificial-intelligence-is-changing-the-construction-industry/
[7]. Clavero, J. 2018, Artificial Intelligence in construction: The Future of Construction https://esub.com/artificial-intelligence-construction-future-construction/,
[8]. Bartneck, C., Lütge, C., Wagner, A., Welsh, S. (2021). Privacy Issues of AI. In: An Introduction to Ethics in Robotics and AI. SpringerBriefs in Ethics. Springer, Cham. https://doi.org/10.1007/978-3-030-51110-4_8

Baby Sharma, et. al. "AI and data privacy in ConTech: A case study of viAct's responsible scenario based AI." *International Journal of Computational Engineering Research (IJCER*), vol. 12, no.2, 2022, pp 01-07.