Student Name: Lingfeng Zhang

Student Number: 300134245

Topic: This government proposes to make Deep Learning the preferred AI technique for all medical decision-making by 2030.

1. Background and Description of the AI in Question:

In the conventional medical system, doctors interpret medical images and help patients to diagnose their diseases according to doctors’ previous knowledge. However, this is quite time-consuming and biased sometimes because of its subjectivity and the complexity of the medical images. Due to the development of artificial neural network, which is conceptually inspired by the human biological nervous system[4], and powerful computing resources, deep learning can be applied in medical images segmentation, classification and recognition applications.

Deep learning is an advanced technology, which can analysis medical images and determine whether the patient suffers from a disease and investigate what kind of disease with which the patient is infected. Deep learning acts as a black box, meaning by given a medical image, pretrained deep learning model can analysis this image with the extremely high accuracy although small errors happen sometimes with no reasons. However, nobody knows how the deep learning model gives this certain response clearly because in my view, deep learning is an optimization problem, the model starts from “guessing” the result randomly until predicting the result with more and more accuracy. There is no special and explicit explanation why deep learning can perform well.

Since deep learning system programmers do not understand the AI system completely, they may not take the correlative responsibility if some errors or mistakes occur in this system. Hence, nobody will be blamed and take the responsibility, then the victim will not be willing to seeing this situation happens.

Currently, many deep learning-based computer vision applications are performing even better than humans, including medical images recognition and other medical analyzing systems. They are able to identify many diseases like cancer in blood and tumors in MRI scans.[4] This is the reason why governments propose to make deep learning as preferred AI technique for all medical decision-making even though the lack of interpretability, explainability and understandability, compared with conventional machine learning, which is a well-illustrated AI technology.

1. Ethical Implications of the Policy Proposal:
2. Fail to illustrate wrong results.

In deep learning, the prediction results are not always 100% accurate and sometimes the deep learning model misclassify the result with no clear reasons. There are some adversarial attacks in the deep learning field, which means a tiny perturbation in the image can destroy deep learning model completely. In this case, deep learning models will be unreliable and nobody can illustrate this phenomenon clearly.

Watson’s descendants will act as the medical experts in the healthcare filed[5] as well as Deep patient[1] is used in real world. Due to the limitation of deep learning, some issues may occur. If the patient is suffering from the disease especially the fatal disease and the artificial intelligent medical decision-making system fails to detect it. The patient will lose the best time to get treatment and this mistake may conduce the death of the patient. On the other side, if a heathy person is falsely detected a seriously disease, the psychological hint probably has a negative effect on the emotional health of this “AI victim”, especially who trust the AI technology completely.

1. Privacy issues

Due to properties of deep learning, training a deep learning model needs a large dataset. If the medical decision-making system is applied totally, a lot of patient’s personal medical information will be stored and used in model training. Since patient’s private information is stored in the database, many hackers may have the ability of stealing these private data or analysis the data source, meaning they can identify individuals given enough data[3], according to pre-trained models. This case may offense people’s privacy, which is privacy issue in AI.

Sometimes patients are unwilling to expose their personal medical records, which will be used in analyzing other patients’ diseases. These patients have the right to protect their individual information and hospitals also have the responsibility to respect patients’ privacy.

Without large amount of dataset to train, deep learning model cannot perform well. Hence, implementing deep learning-based medical decision-making system needs patients’ medical records. So, policies should be proposed to tackle with this privacy issue.

1. Normative Recommendations:
2. Human intervention

In most deep learning systems, interpretability, explainability and understandability is important because fairness may be too abstract to be completely encoded into the system. In addition, even bias, like gender-biased word embedding pattern, may occur sometimes in AI systems because we do not know how deep learning models interpret the relationship between various words detailly. Moreover, if the dataset is unclean, for instance, maybe biased text sources, word embedding will be biased in the high probability. Without human intervention, AI decisions must be understandable for average person to contest. However, the AI system is so complicated that even designers of this system may struggle to isolate the reason for any single action.[1]

In another way, explanation is not always necessary either because there are not significant consequence even though we do not know how the system gets the final result, or because the problem is sufficiently well-studied by AI applications and we can trust the system’s decision even though the system is not 100% accurate. So, not all machine learning systems require interpretability.[2]

Since in the conventional medical decision-making case, even patients do not understand medical details and methodologies. Doctors can diagnose patients’ health condition with the help of deep learning tools and improve the efficiency and productivity. The process is deep learning systems diagnose patients’ health and doctors verify the results. Even though we do not understand how deep learning get the certain result, doctors can play the role of monitor and take the relevant responsibilities.

1. Privacy protection policy.

To protect the personal privacy, The European Union published GDPR policy, which is the abbreviation of General Data Protection Regulation.[3] GDPR is meant to protect the average person. In addition, GDPR will apply to almost all activities when using customers’ personal private information. Another case is WP29, which is similar policy as GDPR. WP29 is a full-blown prohibition of automated decision-making with some limited exceptions.[6]

With these privacy protection policies, the cost of stealing personal private data will be so heavy. It will effectively avoid privacy issues in AI systems.

To be simplification, some patients’ private information which will be identified a certain person can be removed from the dataset, like personal name, social insurance number and so on. In machine learning, these features sometimes are not so important to analyze new patients’ diseases actually.

1. Possible Objection and Responses:

According to the previous normative recommendations, maybe someone will disagree my idea, and say AI prediction algorithms may analysis the information to point out a certain person and these person’s privacy will also occur the risk of exposure.

In my mind, although some algorithms can do so, these predictions are not always 100% accurate and nobody is willing to trust it with 100% certainty. In another word, there is no proof to show that this is the information of this certain person only relying on the machine prediction. In addition, government policy will restrain these evil behaviors because offenders will receive the punishment eventually.

In another way, when the system is collecting the information of patients, it will be granted from the patients’ consents. However, collecting user consent is difficult and managing user consent is not easy.[3] Is this meaning system can use users’ private information without limitation if users grant the privilege to it? This is another ethical issue should be tacked soon.

References:

[1] Knight, Will. "The dark secret at the heart of al." Technology Review 120.3 (2017): 54-61.

[2] Doshi-Velez, Finale, and Been Kim. "Towards a rigorous science of interpretable machine learning." arXiv preprint arXiv:1702.08608 (2017).

[3] Andrew Burt. (2018) How will the GDPR impact machine learning? O'Reilly Media

[4] Razzak M.I., Naz S., Zaib A. (2018) Deep Learning for Medical Image Processing: Overview, Challenges and the Future. In: Dey N., Ashour A., Borra S. (eds) Classification in BioApps. Lecture Notes in Computational Vision and Biomechanics, vol 26. Springer, Cham

[5] Millar, Jason, and Ian Kerr. "Delegation, relinquishment, and responsibility: The prospect of expert robots." Robot Law. Edward Elgar Publishing, 2016.

[6] Andrew Burt, Stuart Shirrell. (2017) Why we're concerned about the WP29's guidelines on machine learning. Privacy Perspectives.