A Utilitarian Analysis of EU's CE Mark for ChestLink: Autonomous Artificial Intelligence for Medical Diagnostic Evaluation

STS Research Paper
Presented to the Faculty of the
School of Engineering and Applied Science
University of Virginia

By

Anna Workman

March 27, 2024

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISOR

Benjamin J. Laugelli, Assistant Professor, Department of Engineering and Society

Introduction

On March 29th, 2022, Oxipit's CEO, Gediminas Peksys, announced that

its technology, ChestLink, had received certification from the European Union (EU) and would become the first Artificial Intelligence (AI) tool to carry out a medical diagnostic evaluation autonomously (Oxipit, 2022a). ChestLink is certified as a Class IIb medical device and its mark is issued for a 'Standalone computer-assisted diagnostic medical device for chest X-ray analysis and reporting.' (Oxipit, 2022b). As of 2023, ChestLink was deployed in a clinical setting in 32 European countries (Oxipit, 2022a).

Artificial intelligence in medical diagnostics is a new territory for regulatory bodies. To date, scholars have acknowledged the importance of ChestLink as the first fully AI medical imaging application to perform diagnostics autonomously (Overman, 2022). However, the analysis of the EU's decision to grant ChestLink certification has focused solely on estimating clinical impact (Plesner et al., 2023). More generally, scholars have begun to ponder the ethical implications of AI in medical diagnostics and the challenges of regulating the development of AI (Karmakar, 2021).

The current understanding of the morality of the EU's decision to grant ChestLink certification fails to holistically evaluate the utility and consequences of adopting this technology in medical practices. By continuing to look at the technology from a purely clinical safety and performance perspective, regulators could miss the vital implications of adopting this novel technology into radiology. The EU's regulatory process for granting ChestLink a CE mark should have considered a wider breadth of the impact on the well-being of all stakeholders.

I will apply utilitarianism, an ethical framework where moral actions are based on maximizing the overall happiness of most people, to evaluate whether the EU should have

awarded ChestLink with certification. By displaying the larger picture of the regulatory decision, I will show that the EU's decision to allow the use of ChestLink in diagnostics was moral. Through the lens of utilitarianism, I will assess the EU's regulation pathway, Oxipit's data on performance, third-party evaluations of the software, studies on patient trust, and interviews with radiologists.

Background

ChestLink is an AI tool that was developed by the company Oxipit. The AI software analyzes chest X-rays. If ChestLink is highly confident that the images feature no abnormalities, it produces an automated report. In any instances where the confidence is not high or there are suspected abnormalities, the image will be sent to a human radiologist for usual reporting. Oxipit claims that radiology departments are understaffed by one-third in developed countries. Additionally, the company cites the World Health Organization statistic which states that "two-thirds of the world population does not have access to diagnostic medical imaging." The goal of ChestLink is to reduce radiologist's workload.

Literature Review

As artificial intelligence technology grows and is applied more to the world of medicine, scholars have begun to examine the importance of ethics, privacy, and security in this field. Thus far, scholars have approached the examination of Oxipit's ChestLink from a purely performance-based assessment (Plesner et al., 2023). Additionally, ChestLink has been discussed in the context of bigger discussions about the regulation of AI in cardiovascular imaging. Scholars have failed to ask questions about the morality of granting ChestLink Class IIb MDR certification.

Studies have been conducted to compare the performance of ChestLink with clinical radiology reports to estimate its clinical impact (Plesner et al., 2023). Scholars characterized the

results of their study by looking at the specificity and sensitivity of the tool and concluded that automated normal chest radiography interpretation could alleviate the workload of radiologists (Plesner et al., 2023). However, they fail to measure the actual implementation of the AI tool and more importantly, the responsibility of regulatory bodies to determine the potential clinical effects of ChestLink beyond performance metrics.

Ernst Wellnhofer argues that AI in cardiovascular imaging raises new challenges for regulatory bodies and that they must develop new controlling techniques and continue ongoing regulatory research (Wellnhofer, 2022). Wellnhofer discusses the pathway that ChestLink took to become the first Class IIb MDR-certified radiological imaging AI tool to be used in a clinic without any involvement from a radiologist (Wellnhofer, 2022). He states that the regulatory approach must be risk-based and resilient with a focus on monitoring unknown emerging risks (Wellnhofer, 2022). Wellnhofer discusses the technical risks of AI in cardiac imaging including input attacks and hidden biases rooted in the data (Wellnhofer, 2022). In such a case, "MLMD should be accurate, robust, trustworthy, transparent, explainable, understandable, and resilient." (Wellnhofer, 2022). The European Commission has created guidelines for trustworthy AI, stating that it must be lawful, ethical, and robust (*A European Approach to Artificial Intelligence*, 2024). Despite these claims, Wellnhofer fails to assess the morality of the EU's decision to grant Oxipit the power to market its device. By addressing these shortcomings and looking at the certification decision through the lens of utilitarianism, I will demonstrate that their decision was moral. A utilitarian analysis will enable me to identify the utility, universalism, and consequences of ChestLink.

Conceptual Framework

To analyze the morality of the EU's decision that enabled the use of ChestLink in clinics, I will be using Jeremy Bentham's ethical framework of utilitarianism. The framework will allow me to take a pragmatic approach to evaluate the ethics of the EU's decision-making and examine the overall impact on the well-being of those being impacted directly, and indirectly by ChestLink. Utilitarianism is built on a class of ethical theories called consequentialism, where the moral judgment of an action is based on its consequences (van de Poel & Royakkers, 2011). The utility principle, also known as the "greatest happiness principle", judges the usefulness of actions by their tendency to invoke pleasure or happiness or prevent pain and unhappiness (van de Poel & Royakkers, 2011). According to van de Poel, utilitarianism is a monistic type of consequentialism. Together, the utility principle and consequentialism intertwine to produce the utilitarian principle whereby all actions shall be judged by the amount of pleasure and pain they bring about (van de Poel & Royakkers, 2011). Utilitarianism concerns itself with values because it is critical to know what is intrinsically good to determine the moral rightness of the action.

To measure the pleasure and pain produced by an action, Bentham created the concept of a moral balance sheet. The moral balance sheet is a cost-benefit analysis table that allows one to calculate the utility of an action (van de Poel & Royakkers, 2011). It uses the criteria of intensity, duration, certainty, remoteness, redundancy, purity, and extent. I will also be using an extension of utilitarianism known as distributive justice, which considers the distribution of costs and benefits. Henry Sidgwick describes moral action as being the one that results in the most equitable distribution of happiness that must be selected from various situations with equal happiness (van de Poel & Royakkers, 2011).

Drawing on utilitarianism in the analysis that follows, I begin by using the utility principle to assess the EU's decision in terms of the consequences it may have on various stakeholders, including patients and radiologists. In the following section, I will apply distributive justice to examine how the benefits and potential harms are distributed among different stakeholders. Utilitarianism, along with the utility principle, and a consideration of distributive justice, provides a nuanced understanding of the ethical implications of ChestLink and a comprehensive and systematic approach to evaluating the morality of the EU's decision.

Analysis

Utility Principle

Based on the utility principle, the EU acted morally by granting ChestLink a CE mark, thus maximizing the happiness of most people. According to scholar Ruut Veenhoven, four qualities of life encompass how humans experience happiness. These qualities are the livability of the environment, life-ability of the person, utility of life, and satisfaction of life (Veenhoven, 2010). The impact of ChestLink is likely to impact the livability of the environment and the life-ability of the person.

Happiness gained from the livability of one's environment stems from having good living conditions, in which the healthcare system plays a huge role. A good healthcare system has good quality healthcare products. ChestLink is a class IIb medical device (*ChestLink*, 2023). Since it poses a medium-high risk to patients, the EU requires that it meets general safety and performance requirements under MDR 2017/745 (*The European Union Medical Device Regulation*, n.d.). To pass this independent assessment from a notified body, Oxipit must implement a Quality Management System (QMS). This requires applying the EN ISO 13485 standard, performing clinical evaluation, post-market surveillance, and post-market clinical

follow-up plans (*The European Union Medical Device Regulation*, n.d.). Public access to approval documents in the EU is extremely limited and, in this case, could not be found. However, the product examination is performed by a notified body that ensures compliance with ChestLink across all requirements. Therefore, users can rest assured that it has passed safety and performance criteria. Having access to and trusting that this technology is safe and effective will enhance the living conditions of people getting chest x-rays and as a result increase happiness.

To further support the argument that ChestLink is augmenting the livability of the environment from a healthcare performance standpoint, I present the following results from a third-party review. The external evaluators obtained chest radiographs from emergency department patients, in-hospital patients, and outpatients in Denmark. After applying ChestLink algorithms and comparing them to Radiologist reports, they concluded that AI outperformed human radiologists (Plesner et al., 2023). ChestLink's sensitivity for abnormal chest radiographs was 99% whereas the clinical radiology report had a sensitivity of 72% ($p < 0.001$) (Plesner et al., 2023). When the sensitivity of ChestLink was >99% for an abnormality, there was no potential harm to autonomous reporting of normal radiographs (Plesner et al., 2023). Among the 1529 chest radiographs, the AI tool only had one false negative that was clinically significant (Plesner et al., 2023). The practicing radiologist also missed the subtle legion in their clinical practice. The very nature of ChestLink provides hospitals with a highly sensitive tool for saving time when generating reports for normal chest X-rays. In this manner, ChestLink provides a more effective method for reviewing images compared to the existing method. The EU's decision to allow the commercialization of ChestLink gives hospitals access to a technology that will improve their healthcare system, thus providing citizens with a more livable environment and increasing happiness.

ChestLink could also improve the landscape of healthcare by increasing the volume of scans that can be evaluated. When the application is highly confident that the image features no abnormalities, it will produce a final study report. This process takes 10-60 seconds when performed by ChestLink (*Radiology, Health AI Register*, 2023). Since the application can always be on, it can analyze vast amounts of data. For perspective, ChestLink has the capability of evaluating 8654 images over 24 hours. Removing the time that radiologists spend looking at normal images will allow them to invest more time in examining complex cases where AI has not evolved enough to perform. This increase in efficiency of the imaging analysis system will allow more patients to be seen, thus bettering the livability of one's environment and overall happiness.

The life-ability of a person refers to the inner life-chance of a person, or in other words, how well-equipped one is to cope with life's problems (Veenhoven, 2010). Being equipped with potentially life-saving tools like ChestLink can help increase the life-ability of people, and therefore their overall happiness. If the technology for ChestLink were to evolve to encompass more assistance in diagnoses, it could save many lives. Currently, ChestLink's algorithm supports 75 different pathologies (*Radiology, Health AI Register*, 2023). One example of where ChestLink has the potential to increase the life-ability of patients and thus overall happiness is in the case of lung cancer. Lung cancer has one of the poorest survival outcomes of all cancers (del Ciello et al., 2017). To properly treat the disease, it must be identified in chest radiographs. In approximately 90% of cases where lung cancer is missed, this fatal miss is attributed to errors in diagnosis on chest radiographs (del Ciello et al., 2017). By allowing the clinical use of ChestLink, the EU has enabled software developers to develop algorithms and apply them in more complex cases where pathologies such as lung cancer can be more accurately detected by

AI than by human radiologists. The opportunity for this technological evolution may lead to an increase in the life-ability of lung cancer patients, or other lung diseases because pathologies could be identified and treated sooner. Family and friends of those patients will also experience greater overall happiness due to satisfaction of care, and longer life expectancies of their loved ones.

Above, I argued that according to the utility principle, the EU's decision to certify ChestLink was a morally sound one that maximized the happiness of the greatest number of people. However, some may argue that the happiness of both patients and doctors could decrease with the use of ChestLink in clinical settings due to a lack of trust. This is a reasonable point of view because trust is a value that shapes the livability of one's environment, and therefore overall happiness. Additionally, scholars have argued that a lack of trust in AI systems is a significant barrier to overcome to successfully implement the technology in healthcare (Asan et al., 2020). An experiment performed by Georgiana Jurvale et al. showed that patients are more likely to trust a human over an algorithm when it comes to their primary diagnosis (Juravle et al., 2020). In a video of Radiologist Dr. Cellini, he expressed a sense of duty to check over all the AI-generated reports not only out of distrust but because if they were wrong, he could be the one held liable for any resultant harm (Cellini, 2022).

I argue that while trust in AI may not be at an optimal level now, it will increase and eventually result in the greatest happiness for most people. Humans' preference towards human doctors for AI diagnosis is likely a learned one, as we have interacted with humans for our whole lives up to this point. Trust in AI diagnosis increases after patients are informed that it has been adopted by others, or through the construction of preference (Juravle et al., 2020). When it comes to ChestLink, the patient is still getting critical doctor-patient interaction time. The AI tool

is only used to generate the radiologist report for normal chest X-rays that have a high confidence interval. Clinicians are held responsible for the harm to a patient if they deviate from the standard care procedure for a given health concern. Analogous to a scenario where a human radiologist writes a normal report, but the patient is still experiencing symptoms, the doctor must use their best judgment to either get additional scans or get the scan double-checked by another radiologist. Trust in AI systems is based on both organizational policies and culture (Asan et al., 2020). Legislation will evolve and adapt to the influx of this new technology. As AI is increasingly used in healthcare, both patients and physicians will continue to gain trust in the technology, which along with the above evidence, will maximize happiness. Thus, according to the utility principle, the EU made a moral decision.

Distributive Justice

Finally, I argue that through the lens of distributive justice, the EU made a moral decision to award ChestLink with a CE mark. Using ChestLink in a clinical setting will result in a more equitable distribution of happiness. ChestLink removes inter-radiologist subjectivity. In a study measuring performance in Chest Radiography, results demonstrated a substantial variability in the performance of radiologists interpreting a standardized set of chest radiographs, the differences between each group of observers were statistically significant ($p < .001$) (Potchen et al., 2000). Authors attribute this interobserver variability and diagnostic accuracy to individual medical training, the image-viewing environment, access to previous radiographs, and/or pertinent clinical information (Potchen et al., 2000). This variability in performance leads to inequitable distribution of access to radiologists with high accuracy rates. Well-trained radiologists are more likely to reside in urban areas and well-funded hospitals. Affluent populations and those with comprehensive insurance coverage are more likely to have access to

these limited services. Meanwhile, ChestLink can be implemented in all medical facilities and regions of the world. There are population and technical factors that determine the accuracy of diagnostic software such as ChestLink (Fanni et al., 2023). While these factors must be monitored, they are more controllable than human accuracy factors and can be implemented on a much larger scale. As a result of granting ChestLink a CE mark, health care systems can evenly distribute the technology and allow equal performance for detecting normal chest x-rays in clinics.

ChestLink provides the opportunity to equitably distribute benefits across all populations. Already, results from an external evaluation have shown that ChestLink's autonomous reporting rate of normal chest radiographs was not significantly different across the four hospitals where images were tested (Park, 2023). Additionally, the EU's certification process involves post-market surveillance and clinical follow-up plans. As AI becomes more prominent in healthcare, there will be more regulations put in place. These regulations will ensure that the algorithms are free from bias and that training data comes from a diverse set of populations. In areas where properly trained radiologists are hard to access and radiologist performance variance is high, ChestLink is a standard tool that applies the same algorithm to all images. Consistency in performance across all populations is critical for distributive justice, and the EU has granted certification to a technology that has proved it is on track in this regard.

Across the world, there is a shortage of radiologists, a solution that is not easy to remedy when it takes 8 years to train a radiologist. Even before the pandemic, in the UK, 99% of radiology departments could not meet the increasing demand for scan and X-ray reporting (*Specialty in Europe*, n.d.). Without timely diagnosis, patients are at risk of not receiving life-saving treatments. These shortages have even greater impacts on countries with

disproportionately high burden rates, increased disease rates, and poor health outcomes. In these environments, radiologists are under pressure to read reports at a faster rate. There is a positive correlation between faster radiologist reading speed and the number of major misses and interpretation errors (Sokolovskaya et al., 2015). Depending on the type of medical institutions and types of cases presented, ChestLink may automate from 15% to 40% of daily reporting (Oxipit, n.d.). As a result, ChestLink offloads the burden to healthcare systems that are suffering from shortages of radiologists. Consequently, happiness is distributed equitably and results in good for most people, thus corroborating the morality of the EU's decision.

Conclusion

Through the lens of utilitarianism, I have argued the morality of the EU's decision to grant ChestLink a CE mark. By applying the utility principle, I demonstrated that allowing ChestLink to be used in clinics will increase the livability of the environment and life-ability of the population, thus increasing overall happiness for the majority. I also considered the distribution of the cost and benefits of ChestLink and claimed that it provides a better opportunity for equity and access to accurate radiology reports.

By evaluating the EU's decision from a utilitarian perspective, I have shown the importance of understanding the consequences for the well-being of all stakeholders. Regulatory bodies must look beyond the normal safety and performance standards for AI in medical diagnostics. They must consider what is best for everyone and ensure that evaluation methods are rigorous. AI tools such as ChestLink will continue to be approved for use in clinics worldwide and we must hold regulatory bodies accountable for making moral decisions while also encouraging legislation to keep up.

References

Asan, O., Bayark, A. E., & Choudhury, A. (2020). Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *Journal of Medical Internet Research, 22(6)*. https://www.jmir.org/2020/6/e15154/

Cellini, M. [Dr. Cellini]. (2022, April 10). *Don't go into radiology - AI is taking over* [Video]. YouTube. https://www.youtube.com/watch?v=Ck9rxqYGpq8

del Ciello, A., Franchi, P., Contegiacomo, A., Cicchetti, G., Bonomo, L., & Larici, A. R. (2017). Missed lung cancer: When, where, and why? *Diagnostic And Interventional Radiology* 23(2): 118–126. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5338577/

European Commission. (2024, February 7). A European approach to artificial intelligence. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

Fanni, S. C., Marcucci, A., Volpi, F., Valentino, S., Neri, E., & Romei, C. (2023). Artificial Intelligence-Based Software with CE Mark for Chest X-ray Interpretation: Opportunities and Challenges. *Diagnostics*, *13*(12), 2020. https://doi.org/10.3390/diagnostics13122020

Juravle, G., Boudouraki, A., Terziyska, M., & Rezlescu, C. (2020). Chapter 14—Trust in artificial intelligence for medical diagnoses. In B. L. Parkin (Ed.), *Progress in Brain Research* (Vol. 253, pp. 263–282). Elsevier. https://doi.org/10.1016/bs.pbr.2020.06.006

Karmakar, S. (2021). Artificial Intelligence: The future of medicine, or an overhyped and dangerous idea? *Irish Journal of Medical Sciences. 191*, 1991–1994.

Overman, D. (2022). Oxipit Awarded CE Mark for Autonomous AI Medical Imaging Application. *AXIS Imaging News.* https://www.proquest.com/openview/ee4a96d82094589ff757f3d25c4cd4b0/1?pq-origsite=gscholar&cbl=2037571

Oxipit. (2022a, March 29). CE mark for first autonomous AI medical imaging application. https://oxipit.ai/news/first-autonomous-ai-medical-imaging-application/

Oxipit. (2022b, April 5). Autonomous AI medical imaging: Understanding ChestLink. https://oxipit.ai/news/autonomous-ai-medical-imaging-understanding-chestlink/

Oxipit. (n.d.). ChestLink—Automated chest x-ray reporting. Retrieved February 28, 2024, from https://oxipit.ai/products/chestlink/

Park, C. M. (2023). AI: Workload Reduction by Autonomous Reporting of Normal Chest Radiographs. *Radiological Society of North America*. https://pubs.rsna.org/doi/10.1148/radiol.230252

Plesner, L. L., Muller, F. C., Nybing, J. D., Laustrup, L. C., Rasmussen, F., Nielsen, O. W., Boesen, M., & Andersen, M. B. (2023). Autonomous Chest Radiograph Reporting Using AI: Estimation of Clinical Impact. *Radiological Society of North America*. https://pubs.rsna.org/doi/10.1148/radiol.222268

Potchen, E. J., Cooper, T. G., Sierra, A. E., Aben, G. R., Potchen, M. J., Potter, M. G., & Siebert, J. E. (2000). Measuring Performance in Chest Radiography. *Radiological Society of North America.* https://pubs.rsna.org/doi/abs/10.1148/radiology.217.2.r00nv14456?journalCode=radiology

Radiology, Health AI Register. (2023, November 23). ChestLink. *Health AI Register*. https://www.healthairegister.com/products/oxipit-chestlink/

Sokolovskaya, E., Shinde, T., Ruchman, R. B., Kwak, A. J., Lu, S., Shariff, Y. K., Wiggins, E. F., & Talangbayan, L. (2015). The Effect of Faster Reporting Speed for Imaging Studies on the Number of Misses and Interpretation Errors: A Pilot Study. *Journal of the*

American College of Radiology: JACR, *12*(7), 683–688. https://doi.org/10.1016/j.jacr.2015.03.040

Specialty in Europe. (n.d.). European Union of Medical Specialists. Retrieved February 28, 2024, from https://www.uemsradiology.eu/specialty-in-europe/

The European Union Medical Device Regulation. (n.d.). Retrieved February 28, 2024, from https://eumdr.com/manufacturers/

van de Poel, I., & Royakkers, L. (2011). *Ethics, Technology, and Engineering*. John Wiley & Sons, Ltd.

Veenhoven, R. (2010). Greater Happiness for a Greater Number. *Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht. 11, 605–629.*

Wellnhofer, E. (2022). Real-World and Regulatory Perspectives of Artificial Intelligence in Cardiovascular Imaging. *Frontiers in Cardiovascular Medicine, Volume 9.* https://www.frontiersin.org/articles/10.3389/fcvm.2022.890809/full