# Ethics in Lung Cancer Clinical Diagnosis

Lab IACD - Group 8

October 28, 2024

Along with making a **predictive model** comes the responsibility of managing and processing a dataset. In the case of this project, said data involves **CT Scans** of real people and information on nodules (that could be cancerous) within their lungs. Because of this fact, we must ask ourselves: Are we dealing with sensitive data? Is the dataset anonymized? What measures should we take to protect said data?

When handling this kind of information, we must be careful about what is used to build the model and what should be scrapped, before even thinking about analysing it. **PI** (Personal Information) should all be removed from the dataset to prevent a third party from identifying any of the subjects at hand. All other data contained in it has to be anonymized, removing the possibility of cross-referencing the data with another dataset.

Upon analysing it, we can infer that all the **PI** of the patients, as well as all the **PHI** (Protected Health Information) in the DICOM files, were removed before we got access to it. The annotations provided with each nodule were also properly checked, so as to not show any specific marks from any of the radiologists involved in the collection of data.

However, this dataset is still at a risk that can't be prevented: it is **pseudo-anonymous**. We may try to protect all the information in it, but, because we're dealing with images, it is always possible to cross-reference them with any hospital's public datasets. Although this is a liability, there is no way to anonymize the images without losing their data in the process.

In the future, the dataset may have to go through changes to accommodate new **laws and regulations**. Making sure that the data is regularly checked and complies with Privacy Law is vital. The classification models created using it must also be in accordance with these laws and the decisions made by them should be easily explained. This way, the doctors who end up using them can know when the model is best used and what it can/can't do.

At the moment, the regulatory framework for AI-based healthcare models, such as those developed using this database, treats these systems as medical devices. In the U.S., the **FDA** (Food and Drug Administration) classifies them as **SaMD** (Software as a Medical Device), whereas in the EU, the **MDR** (Medical Devices Regulation) oversees similar standards, focusing on safety, performance, and transparency. Other regulations, like the **Artificial Intelligence Act** that was implemented this year, aim to classify AI systems by risk, imposing stricter requirements for high-risk applications like diagnostic tools. As these frameworks evolve, developers will need to address ethical concerns around bias, privacy, and accountability, ensuring that AI models are fair, transparent, and prioritize patient safety.

## Bibliography

https://aapm.onlinelibrary.wiley.com/doi/full/10.1118/1.3528204