



Enforce GDPR Compliance in Machine Learning

- Summer Internship at Radix -

About Radix

Radix is an Artificial Intelligence solutions provider based in Brussels. Together with our clients, we deliver AI solutions that have a positive impact on the world.

Our mission is to help our clients grow and to improve people's lives. To deliver on our promise, we focus on the combination of three elements:

- Business: creating high value for our clients, their employees and society as a whole
- People: delivering technology that can truly assist people, not replace them
- AI: automating intelligence to achieve the highest impact

Founded in January 2018, Radix is now a team of 25 people that continues to grow. We work with ambitious clients like GSK, Brussels Airport, Atlas Copco, Flanders Investment and Trade, VDAB, Belga, Macadam, House of HR and more.

Visit our website <https://radix.ai> for more information.



The internship

As of this year, Radix will host several summer internships. Such an internship shows you, among other things, how your learning material is applied in the field and what the best practices are when working on a project, this while being supervised by field experts. Aside from that, it allows you to get a glance of how a growing startup operates from the inside and how other fields such as *Marketing* and *Sales* contribute to this.

About the project

Over the years, protection of personal data has taken the spotlight within the technology industry. One of the most prominent legislations regarding the protection of personal information is the GDPR. However, enforcing GDPR compliance isn't as straightforward when working with big data. This project investigates the difficulties that arise regarding data protection when handling large amounts of textual information, and how these are handled.

What you will learn

During this internship, you will learn how to detect personal data within a large text corpus, and how to censor this information properly. In order to do so, you'll create a predictive model that detects such fields based by taking contextual data into account. The goal is to create a fully automated pipeline that takes as inputs these texts, and outputs the detected personal data as well as a censored format of the input-document.

Your profile

- You are in your third Bachelor or first Master in Computer Science or a related field
- You are fluent in English, both written and verbally
- You have strong analytical skills and are familiar with the basics of machine learning
- You know your way around in Python
- Familiarity with statistical Python tools such as Spacy, TensorFlow, or Torch is a plus
- Basic knowledge of GDPR legislation required
- Experience within NLP is a plus
- You are eager to learn and are open to be challenged