Setup NL AIC Working group
“Teaching Responsible AI (TRAI)”
v1.1, April 2023

Introduction

This document outlines the motivation, goals and positioning of the working group Teaching Responsible AI (TRAI) of the NL AIC (Nederlandse AI Coalitie). The TRAI working group was initiated by the following people:

- Francien Dechesne (UL)
- Maaike Harbers (HR)
- Marieke Peeters (Mooncake-AI)
- Birna van Riemsdijk, chair (UT)
- Pascal Wiggers (HvA)

TRAI has been initiated as part of the Participative and Constructive Ethics (PACE) working group of the Human Centric AI building block of the NL AIC. Its theme regarding education moreover has links with the Human Capital building block, in particular the working group Training and Education. We discuss these connections in more detail below.

Motivation

Currently, many organizations are adopting artificial intelligence (AI), and want their organizations to become more AI-driven[1, 2]. Most of them are already taking steps towards increasing their AI maturity. However, applying AI technology requires more than technological expertise.

By now, companies and public institutions that have developed their AI maturity to substantial levels will mostly harness an abundance of technical experience. But at the same time, their technical workforce often lacks the required domain knowledge and experience to understand how the business and other stakeholders could either benefit from or risk being harmed by AI-driven innovation. In order to apply AI technologies in an effective and responsible manner, organizations must develop a solid understanding of the intended and established (ethical, legal, societal, and environmental) impact of the solutions they are implementing within their organization as well as within their broader field of work. And so the people working in technical

[1] AiNed programme, door NL AIC (2020)
[2] Opgave AI. De nieuwe systeemtechnologie | Rapport, door WRR (2021)

departments must develop knowledge and skills needed to understand and mitigate the particular risks of applied AI within their company's specific sector (e.g. unfairness and unequal treatment, opaqueness / black boxes, unreproducible results, etc.).

The combination of deep technical knowledge and deep domain knowledge is crucial in order to responsibly and sensibly apply AI[3]. The observed gap between the domain experts on the one hand, and the technical experts on the other, needs to be closed[4]. To do so, knowledge building must be accomplished on two sides[5, 6, 7, 8]: On the one hand, people studying or working within the application domain must learn more about the potential opportunities and risks of AI. On the other hand, students and professionals already working within the broader field of AI[9] need to learn more about the particular application domain in which they apply their skills. In addition, they must learn how to create AI products that not only provide value for the business, customers, employees, or society in general - depending on what values the company or organization pursues[10], but also respect the boundaries and constraints placed on such products and applications.

The current working group aims to tackle the second type of knowledge building: aiding students and professionals within the field of AI to learn more about everything related to the responsible (and sensible) application of AI within a company or organization. The goal is to increase the awareness, knowledge and skills of both students and companies about how to apply AI responsibly, sensibly, and to create value for the stakeholders, as well as when *not* to use AI. Therefore, to establish true impact, we aim to bring together educational institutes and the organizations where students end up working after graduation. By involving both the educational institutes and those organizations, we can ensure that the educational programs teach students the skills needed to help innovate the companies and public institutions of tomorrow, and that these organizations are ready to embrace and make good use of the knowledge brought in by the newly graduated.

Vision and ambition

The fast pace with which AI has moved from an academic discipline to an applicable technology thus requires professionals with new skills. As the WRR noted in their recent report[10], AI is not just a technology, but will fundamentally change society.

---

[3] ELSA Labs for Human Centric Innovation in AI, door NL AIC (2021)
[4] Manifest Mensgerichte Artificiële Intelligentie, door NL AIC (2020)
[5] AI is Mensenwerk - NL AIC Human Capital-subwerkgroep Scholing en Onderwijs, door NL AIC (2020)
[6] Hoe kunstmatige intelligentie artsen in de praktijk kan ondersteunen, door Rathenau (2021)
[7] We denken te snel dat een meekijkende computer het beter kan, door Rathenau (2021)
[8] Grip op algoritmische besluitvorming bij de overheid, door Rathenau (2021)
[9] Examples are scientists, developers, engineers, researchers, architects, designers, consultants, managers, etc.
[10] Opgave AI. De nieuwe systeemtechnologie | Rapport, door WRR (2021)

As a consequence, there is a need for professionals who understand how to harness the full potential of AI in practice as well as how technology should be embedded in context with an eye for wanted and unwanted consequences of the technology.

Therefore, reflection on the broad impact of AI and on its ethical, legal and societal aspects should be an integral part of any Artificial Intelligence or Data Science curriculum or module. As any decision in a design or development process has ethical consequences and is subject to legal boundaries, these aspects should ideally not be taught in separate courses, but rather run throughout the curriculum. As suggested by (Miller, 1988)[11] technical issues are best understood in their social context, and the societal aspects of computing are best understood in the context of the underlying technical realization.

Educating young professionals to reflect on the impact of technology is only useful if the organizations they will work for welcome such knowledge and provide room for ethical discussion. This also requires education of management. The working group would like to collaborate with organizations and educational institutes that are open to Responsible AI to develop educational material that fits the needs of these organizations and inspires others to adopt a Responsible AI approach.

Goals

The goal of this working group is thus to *bring together TRAI stakeholders who are interested in developing materials, practices and methods of teaching and adopting Responsible AI*[12]. Stakeholders can for example be teachers and students of higher education (HBO/WO) in Computer Science, AI and related disciplines; upper management, employees and talent development professionals of Dutch institutions and companies that use AI systems and develop AI policies and strategy; as well as other beneficiaries. Bringing these groups together will facilitate achieving the following specific aims:

- *Exchanging expertise:* teaching and adopting Responsible AI is challenging due to the inherent multidisciplinary nature of the subject. Responsible AI requires on the one hand deep technical knowledge about AI methods and techniques, and on the other hand insights into the way AI is and could be embedded into organizations and society and associated ethical, legal and societal consequences and risks, combined with an understanding of how one affects the other. This means that engineering skills need to be integrated with an understanding of how technical choices can affect people and society, and ways of engaging with the socio-technical nature of AI systems. Moreover, skills for multidisciplinary collaboration are essential for obtaining such understanding in a concrete work context where expertise of different aspects of the problem often resides

[11] Keith Miller (1988) Integrating Computer Ethics into the Computer Science Curriculum, Computer Science Education, 1:1, 37-52, DOI: 10.1080/0899340880010104

[12] We use the term Responsible AI as a broad term referring to the importance of accounting for ethical, legal, societal, environmental and human-centered aspects when developing AI systems.

with different parts of an organization. Through this working group we bring together teaching practitioners, students and industry and institutional stakeholders of different fields to allow for sharing of expertise, methods, pitfalls, best practices and practical cases in teaching and adopting Responsible AI. In this way we will be able to accelerate the integration of responsible aspects of AI into our curricula and engineering practices, and professionalize our methods and approaches for teaching and adopting Responsible AI.

- *Community building:* besides exchanging expertise on the content and methods of teaching and adopting Responsible AI, we furthermore aim to create a community through which one can meet others with an interest in the topic and share experiences. The community can also provide support and help for those new to teaching Responsible AI in getting started and finding the necessary resources, thereby lowering the threshold of entering the area.
- *Agenda setting:* although the importance of Responsible AI is more and more recognized, the traditional separation of engineering and social sciences still informs much of our current teaching and engineering practices. By coming together through this working group, we will increase our impact in putting this topic on the agenda of educational institutions and industry. It will allow us to create a better connection between what is taught and what is practiced by, on the one hand, bringing in the visions, hesitations, experiences, obstacles, and so on from engineering practices into our educational programs, while on the other hand sharing the latest scientific insights with industry. Finally, by bringing together a TRAI community we will be able to identify gaps and opportunities across organizations and sectors, e.g. for creating shared educational modules. We will leverage the platform and context provided by the NL AIC to achieve the envisioned impact, while facilitating inclusion of other parties.

Related initiatives

There are several initiatives that relate to the proposed working group, but currently, in the NL AIC, there is no working group that specifically focuses on teaching responsible AI to IT professionals. Below we also mention a number of related international initiatives on teaching Responsible Computer Science, but to the best of our knowledge these do not focus specifically on AI. These related initiatives do show that the topic is considered highly relevant, both nationally and internationally, and thereby support the initiative of having a dedicated working group on this topic at the NL AIC.

Within the NL AIC, a first related initiative is the NL AIC building block ‘Human Capital’ which focuses on education of AI in general. This involves vocational and higher education AI programs in the Netherlands (for an overview see their 2021 report ‘Inventarisatie AI-opleidingen en good practices’), as well as, AI education for non-AI professionals. Though *responsible* AI is touched upon by some of the programs, it is not the focus of this working group. A second related initiative in the NL AIC involves the building block ‘Human-centered AI’, focusing on social/ societal, ethical and legal aspects (ELSA) of AI. Members of this building

block developed the concept of ELSA labs, which are currently founded throughout the Netherlands. The building block has also developed and currently promotes an approach of guidance ethics (begeleidingsethiek) for the development and deployment of AI. This building block thus centers on the ELSA/responsible aspects of AI, but it does not focus on teaching.

Outside of the NL AIC, there are many national and international projects on teaching Responsible CS & AI. In the Netherlands, many universities offer courses on (topics related to) responsible AI (see, e.g., the aforementioned report 'Inventarisatie AI-opleidingen en good practices'). Internationally, examples involve the Embedded EthiCS (2019) project at Harvard CS, which proposes to embed ethical aspects throughout the CS curriculum as an inherent part of the technical courses; the Ethics4EU (2019) project, which develops best practices and learning resources for integrating ethical aspects in CS study programs for a wide range of CS topics; and the Responsible Computer Science Challenge (2018) launched by Mozilla, an initiative that funds projects on embedding ethics into undergraduate computer science education. These initiatives show the importance of the topic, and thereby show the importance of a dedicated working group on this topic within the NL AIC.