TO BE fAIr: ETHICAL AND FAIR APPLICATION OF ARTIFICIAL INTELLIGENCE IN VIRTUAL LABORATORIES

F. Caccavale [1]
Process and Systems Engineering Centre (PROSYS),
Department of Chemical and Biochemical Engineering,
Technical University of Denmark, Building 228A, 2800
Kgs. Lyngby, Denmark
ORCID: 0000-0003-0795-5988

C. L. Gargalo
Process and Systems Engineering Centre (PROSYS),
Department of Chemical and Biochemical Engineering,
Technical University of Denmark, Building 228A, 2800
Kgs. Lyngby, Denmark
ORCID: 0000-0002-8740-9591

K. V. Gernaey
Process and Systems Engineering Centre (PROSYS),
Department of Chemical and Biochemical Engineering,
Technical University of Denmark, Building 228A, 2800
Kgs. Lyngby, Denmark
ORCID: 0000-0002-0364-1773

U. Krühne
Process and Systems Engineering Centre (PROSYS),
Department of Chemical and Biochemical Engineering,
Technical University of Denmark, Building 228A, 2800
Kgs. Lyngby, Denmark
ORCID: 0000-0001-7774-7442

Conference Key Areas: *Artificial Intelligence in Education, Ethics in Engineering Education*
Keywords: *Digital Education, Virtual Labs, Ethics, Artificial Intelligence*

[1] *Corresponding Author*
F. Caccavale
fiacac@kt.dtu.dk

doi:10.5821/conference-9788412322262.1398

ABSTRACT

In 1984, the film “The Terminator” predicted that a hostile Artificial Intelligence (AI) will threaten to extinguish humankind by 2029. Even though the real present is quite far from this post-apocalyptic scenario where AI rebels against its creator, a growing concern about the lack of ethical considerations in the use of AI is rapidly spreading, leading to the current “ethics crisis”. The lack of clear regulations is even more alarming considering that AI is becoming an integral part of new educational platforms. This follows the wave of digital transformation mainly induced by the Fourth Industrial Revolution, with advances in digitalization strategies, and the COVID-19 crisis, which forced education institutions worldwide to switch to e-learning. The appeal of AI is its potential to answer the needs of both educators and learners. For example, it can provide help grading assignments, enable tutoring opportunities, develop smart content, personalize and ultimately boost on-line learning. Although the “AI revolution” has great potential to improve and boost digital education, there are no clear regulations in place to ensure an ethical and fair use of AI. Therefore, this work aims to provide a comprehensive overview of the current concerns regarding fairness, accountability, transparency and ethics in AI applied to education, with specific focus on virtual laboratories. The main aspects that this work aims to discuss, and provide possible suggestions for, are: (i) ethical concerns, fairness, bias, equity, and inclusion; (ii) data transparency and digital rights, including data availability, collection, and protection; and, (iii) collaborative approach between disciplines.

1 INTRODUCTION

1.1 Virtual Laboratories and the introduction of Artificial Intelligence

Recent years have seen an increased focus on sustainable development, which is defined by UNESCO as a *“development that meets the needs of the present without compromising the ability of future generations to meet their own needs”*[2]. One of the 17 core objectives presented in the SDG initiative is to improve the quality of education. Some of the proposed goals are to help countries in implementing innovative and context-appropriate solutions. This aims to provide education remotely, leveraging hi-tech, low-tech, and no-tech approaches, as well as to seek equitable solutions and universal access. In these times of educational disruption[3], where, according to UNESCO, never before have so many students been out of school, drastic changes in education have been accelerated, resulting in significant advances in adopting digitalization strategies [1, 2]. In this scenario, virtual laboratories (VL) offer a solution to provide inclusive learning opportunities [1].

[2] UNESCO Sustainable Development Goals, https://www.un.org/sustainabledevelopment/education/ (accessed April 29, 2022).

[3] The United Nations Educational, S. and C. O. COVID-19 Impact on Education (UNESCO) The United Nations Educational, Scientific and Cultural Organization, https://unesdoc.unesco.org/ark:/48223/pf0000380398 (accessed April 29, 2022).

In the field of Chemical and Biochemical Engineering, for example, educational process computer-aided tools are becoming widespread and are proving their relevance. Simulators, such as ASPEN Hysys[4] or SuperPro[5], were already fundamental tools in the teaching of Process Systems Engineering (PSE), but recent events have contributed to expanding their application.

Differentiated and individualized learning has been a priority for educators across domains for years [3]. Educators are also aware that one of the most effective ways of learning is to receive immediate and continuous feedback [4, 5]. This prevents small errors and misunderstandings from propagating. Educational tools, thanks to the recent advances in AI methods, are trying to replicate the same process. This is done by using algorithms to understand the level and preparation of the students and adapting the learnings and exercises accordingly [6]. It follows that Artificial intelligence (AI) is becoming an integral part of new educational platforms, since AI has the potential to answer the needs of both educators and learners by providing help in grading the assignments, enabling tutoring opportunities, developing smart content, or/and personalizing the learning. For example, Lu et al. (2021) introduced RadarMath [6], an intelligent tutoring system to support intelligent and personalized learning for math education. Another solution is provided by Carnegie Learning, developing 'Mika'[6], a tool that uses cognitive science and AI methods to create personalized lesson plans and tutoring that adjusts based on student's performance, and it also provides real-time feedback [7].

1.2 Ethical concerns

As reported in the EU guidelines on ethics in artificial intelligence document[7], the core principle of the EU guidelines is that the EU must develop a 'human-centric' approach to AI that is respectful of European values and principles.

> *"The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union, all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoys a unique and inalienable moral status."*

Studies [8] researching the awareness of ethical issues in e-learning systems focus on how to prevent students from behaving unethically (e.g., copying assignments from their classmates) or assessing their knowledge of what ethical behavior is. Preventing

[4] AspenTech, "Aspen Hysys", 2017. https://www.aspentech.com/en/products/engineering/aspen-hysys (accessed April 29, 2022).

[5] Intelligen Inc., "SuperPro Designer", 2017. https://www.intelligen.com/products/superpro-overview/ (accessed April 29, 2022).

[6] Carnegielearning.com. (2017). https://www.carnegielearning.com/products/softwareplatform/mathia-learningsoftware/ (accessed April 29, 2022).

[7] EU guidelines on ethics in artificial intelligence: Context and implementation (2019). https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf (accessed April 29, 2022).

this kind of behavior is imperative and should not be overlooked in designing an educational on-line tool, since these platforms might be used for grading assignments; therefore, preventing unethical behavior from the students' perspective is a very crucial issue. However, this work aims to focus on safeguarding students and their digital rights through the investigation of ethical behavior in virtual laboratories.

The focus on ethics and fairness in Artificial Intelligence (AI) has gained traction in recent years, as demonstrated by the establishment of various Conferences focusing on ethics in AI. Among those, the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)[8] had its first event dating back to 2018, and the AAAI/ACM Conference on Artificial Intelligence, Ethics and Society[9] will have its fifth edition in 2022.

This recent trend sprang due to the fact that, although AI has proved to be beneficial in many aspects, standing out for the accuracy of predictions and the ability to solve complex problems, it can pose risks to the right of personal data protection and privacy, as well as discrimination, as highlighted by EU guidelines on ethics in artificial intelligence document[10], published in 2019. The EU ethics guidelines provide a good overview of the possible ethical concerns surrounding AI and the principles that should be followed to ensure its ethical applications. On the other hand, these guidelines are non-binding, although concerns over the lack of regulatory oversight to support their implementation have been raised.

On 24 November 2021, the Recommendation on the Ethics of Artificial Intelligence[11] was adopted by UNESCO's General Conference. This work builds on the preliminary study on the ethics of artificial intelligence[12] of UNESCO's World Commission on the Ethics of Scientific Knowledge and Technology (COMEST)[13]. This study emphasizes that currently, no global instrument covers all the fields that guide the development and application of AI in a human-centered approach.

We believe that assessing possible ethical issues regarding how data is collected, stored, and used in modelling when designing a virtual environment is fundamental. Data is currently considered a very valuable resource, which could provide extensive knowledge about a person's habits, preferences, and skills, and could therefore be used, if not regulated, for economic gains outside its primary scope. Therefore, this study aims to build on the current knowledge on ethical applications of AI in

[8] ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) https://facctconference.org/index.html (accessed April 29, 2022).

[9] AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society. https://www.aies-conference.com/2022/ (accessed April 29, 2022).

[10] EU guidelines on ethics in artificial intelligence: Context and implementation (2019). https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf (accessed April 29, 2022).

[11] Recommendation on the ethics of artificial intelligence. https://en.unesco.org/artificial-intelligence/ethics#recommendation (accessed April 29, 2022).

[12] Preliminary study on the Ethics of Artificial Intelligence (2019). https://unesdoc.unesco.org/ark:/48223/pf0000367823 (accessed April 29, 2022).

[13] World Commission on the Ethics of Scientific Knowledge and Technology (COMEST). https://en.unesco.org/themes/ethics-science-and-technology/comest (accessed April 29, 2022).

educational platforms and provide cause for reflection on the multiple possible ethical concerns that introducing data-driven solutions to a virtual laboratory might raise. Our objective is that, by highlighting these aspects, more researchers and practitioners working on Engineering Education might contribute with other reflections and resources in the future, to ensure a more ethical and safe virtual experience for learners, where every aspect of the "digital contract" is transparent.

2 BACKGROUND

This work aims to assess the impact that introducing AI can have on Virtual laboratories from an ethical perspective. In order to do that, we focus on the following clusters of aspects to ensure ethical and fair treatment of data: (i) fairness and bias; (ii) equity and inclusion; (iii) data transparency and digital rights; and (iv) collaborative approach between disciplines.

2.1 Fairness and bias

Fairness in AI refers to the attempt to correct algorithm bias in algorithmic decision making models. These models, if trained on biased data, can fail in a multitude of scenarios. This problem is particularly relevant when the outcome can affect people's lives, leading to cases of sexual hiring, racial bias in the justice system, face recognition issues, predicting car robberies, used by the police in Belgium, or profiling unemployed people, as used in Poland[14].

Virtual Laboratories should refrain from using biased data for their predictions and employ effort and resources in investigating this problem.

We provide a possible scenario to clarify this concept. Assume that we created an online platform to teach bioprocesses to students. The platform contains a section where the students can practice their knowledge through practical exercises that become increasingly more difficult. The recommendation of exercises is powered by a data-driven model. All students start with a few default exercises, necessary to assess their preparation and confidence. Afterward, the model needs to predict the appropriate exercises to recommend to the students. These exercises have a dual objective: to maximize both their learning curve and engagement, since if the exercises recommended are too easy, the students might get bored and not visit the platform anymore. In order to maximize the likelihood of recommending the most appropriate exercise, we train the model on personal information about the learners, such as gender (which, per se, already entails problematic assumptions), age, education, and country of origin. Furthermore, let's assume that the platform is relatively new and therefore we do not have much data, but from the few examples we gathered, it seems that learners that selected "male" as their gender, perform worse than others. If we

[14] EU guidelines on ethics in artificial intelligence: Context and implementation (2019). https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf (accessed April 29, 2022).

feed this information to train the model, the model will be biased and will learn that new students that selected “male” as their gender, would be less trained on the material, and therefore should be recommended easier exercises.

This example shows that it is very important to assess the fairness of the data available before using it to make predictions, which will be inherently biased otherwise. Multiple metrics have been proposed to assess a model’s fairness, distinguished into procedural fairness, which focuses on the decision process, outcome fairness, a consequential approach focusing on the end result, and dynamic fairness, i.e., integrating formal fairness metrics with ethical principles.

2.2 Equity and inclusion

A well-designed VL should enhance culturally inclusive learning [9]. When designing a virtual laboratory, it is important to reflect on how inclusive the content is. If a VL is accessible worldwide, the cultural diversity of the users should be taken into account. For example, some common sayings or metaphors might be difficult to understand or offend the users’ sensibility.

Moreover, a good e-learning platform should account for individual learners’ differences and cater to their learning abilities. This could become an issue of inclusion since it would be unjust to either make the material more accessible for slower learning students or too challenging to benefit the quickest ones.

2.3 Data transparency and digital rights

Before collecting and storing data in virtual applications, it is essential to establish a clear scope and purpose for data collection and analysis. Virtual laboratories should aim to create a feeling of trust in the user, by being clear and transparent on the nature and extent of the data gathered, and what kind and how much of their data is being stored and used [9]. The data collection needs to follow the Principles of GDPR[15]. Moreover, users should be ensured that their data is secure and there will be no breaches in computer ethics. This is particularly important since it often happens that websites are hacked and data accessed. Therefore, gaining the students’ formal consent on how the data is stored and used is very important.

This “transparency contract” also covers AI applications. The user should be informed about what data is fed into data-driven models to improve their experience on the platform and they should have power over their preferences, such as deciding whether to stop allowing their personal data to be used for training models.

[15] General Data Protection Regulation GDPR. https://gdpr-info.eu/pdf (accessed April 29, 2022).

2.4 Collaborative approach

We believe that engineers working on the design and implementation of virtual laboratories should be acquainted with ethical principles and that ethics should be taught in computer science degrees. However, since these issues with ethics would likely require the cooperation of stakeholders of varying expertise in business, law and other disciplines, regulations must be established and enforced at a higher level [10]. The ethics crisis does not only affect engineers and, therefore, should not be solved only by technical people but as a collaborative effort among different disciplines.

As argued by [10], solutions and guidelines to address the ethics crisis cannot be siloed, i.e., each discipline working on a solution without taking into account the point of view of others. This can be exemplified by analyzing the lack of joint outputs and cross-citations among different fields, for example, between articles in computer science journals and articles in the humanistic social science field. Echoing [10], we argue that the best way to address this issue, which should be perceived as a transversal problem that involves theories and methods across several disciplines, is through a collaborative effort between disciplines.

In summary, this study aims to address ethical concerns due to the introduction of AI in virtual laboratories. We aim to build on established knowledge and inspire researchers to contribute to the discussion. In our analysis, we focus on different aspects, such as the evaluation of fairness and bias, equity and inclusion, the necessity for data transparency and privacy protection, and the need for a collaborative effort across different disciplines to establish clear regulations.

3 RECOMMENDED FRAMEWORK AND FUTURE DIRECTIONS

We recommend following these four checkmarks in order to assess the ethical application of AI to virtual laboratories: (i) clarify scope and purpose; (ii) ensure transparency; (iii) protect users' privacy; (iv) help students – the ultimate goal.

(i) Defining a clear scope and purpose for data collection and analysis allows for setting ethical boundaries. Before starting to collect users' data, it is important to be clear on the scope and purpose of the collected data. It is important to reflect on whether all the data points collected are necessary, and establish a plan for exactly how the data will be used. If the scope changes, a new assessment will be necessary.

(ii) Once the data is collected and safely stored, it is important to assess its quality and fairness, before using it to train data-driven models. Assessment of fairness can be performed through various fairness metrics, defined as formal criteria to score fairness. This will build control into the design of the implemented algorithms. Establish guidelines on how to assess whether the data contains bias and remove the biased content from the data. Also, investigate the transparency and understanding of the algorithm: what the prediction is based on, the features that

lead the model to that specific prediction, and consider probing the model to understand what the model knows about the user. Moreover, consider the whole process in a holistic way, rather than analyzing small pieces of the process as a silo.

(iii) Regarding transparency and consent, learners should be aware of how and why their data is being collected and used. This process should be clearly described and shared with the users. Consent must be established before data is collected. Also, consider and clearly communicate whether the students can decide to revise their selected preferences on data sharing at a later stage.

In order to protect the students' personal data, access to the data should be restricted to only the necessary fields, and not all stakeholders working on the VL need access to all data. For example, access to credit card information, passwords, and similar field should be restricted. The available information should also be assessed carefully when selecting the parameters to be used in training a data-driven model.

Finally, we aim to highlight that educational platforms should have the purpose to help students. The ultimate goal of a virtual laboratory is to help people learn, so the solutions implemented should be directed towards this goal. We also argue that platforms should provide the possibility for students to freely share their ethical concerns, ask questions, and in return receive transparent elucidations on their requests.

4 SUMMARY

This study aims to address possible ethical concerns arising from the introduction of Artificial Intelligence in virtual laboratories. Building on established knowledge and regulations, we provide a checklist of aspects to consider when evaluating how to ethically incorporate AI models into an on-line educational platform. Specifically, we focus on fairness and bias, equity and inclusion, the necessity for data transparency, and privacy protection. Moreover, we address the need for a collaborative effort across different disciplines in establishing clear regulations, with the final goal of safeguarding students and their digital rights.

Our objective is to spur researchers and practitioners working on Engineering Education to contribute with reflections and resources to ensure a more ethical and safe virtual experience for learners, where every aspect of the "digital contract" is transparent.

REFERENCES

[1] Gargalo, C. L.,, Heras, S. C. de las, Jones, M. N., Udugama, I.,; Mansouri, S. S., Krühne, U., Gernaey, K. V. (2020), Towards the Development of Digital Twins for the Bio-Manufacturing Industry; Adv Biochem Eng Biotechnol. https://doi.org/https://doi.org/10.1007/10_2020_142.

[2] Narayanan, H., Luna, M. F., von Stosch, M., Cruz Bournazou, M. N., Polotti, G., Morbidelli, M., Butté, A., Sokolov, M. (2020), Bioprocessing in the Digital Age: The Role of Process Models. Biotechnol. J. 15 (1), 1–10. https://doi.org/10.1002/biot.201900172.

[3] Landrum, T. J., McDuffie, K. A. (2010), Learning Styles in the Age of Differentiated Instruction. Exceptionality, 18 (1), 6–17.

[4] Laurillard, D. (2009), The Pedagogical Challenges to Collaborative Technologies. Int. J. Comput. Collab. Learn. 4 (1), 5–20. https://doi.org/10.1007/s11412-008-9056-2.

[5] Hattie, J., Timperley, H. (2007), The Power of Feedback. Rev. Educ. Res., 77 (1), 81–112. https://doi.org/10.3102/003465430298487.

[6] Lu, Y., Pian, Y., Chen, P., Meng, Q., & Cao, Y. (2021), RadarMath: An Intelligent Tutoring System for Math Education. Proceedings of the AAAI Conference on Artificial Intelligence, 35 (18), 16087-16090. https://ojs.aaai.org/index.php/AAAI/article/view/18020

[7] Tan, J. (2018), Discovering Metacognitive Approaches to Personal Finance with Intelligent Tutoring Systems. Georgia Institute of Technology.

[8] Almseidein T, Mahasneh O. (2020), Awareness of ethical issues when using an e-learning system. International Journal of Advanced Computer Science and Applications. 11 (1), 128-31.

[9] Anderson B, Simpson M. (2007), Ethical issues in online education. Open Learning: The Journal of Open, Distance and e-Learning. 22 (2), 129-138. https://www.tandfonline.com/doi/full/10.1080/02680510701306673

[10] Raji ID, Scheuerman MK, Amironesei R. (2021), You Can't Sit With Us: Exclusionary Pedagogy in AI Ethics Education. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAcct '21), 515-525. https://dl.acm.org/doi/epdf/10.1145/3442188.3445914