Global trends in scientific debates on trustworthy and ethical Artificial Intelligence and Education

Christian M. Stracke [1] [0000-0001-9656-8298], Irene-Angelica Chounta [2] [0000-0001-9159-0664], and Wayne Holmes [3] [0000-0002-8352-1594]

[1] University of Bonn, Germany, [2] University of Duisburg-Essen, Germany, [3] University College London, UK
stracke@uni-bonn.de, irene-angelica.chounta@uni-due.de, wayne.holmes@ucl.ac.uk

Abstract. This paper presents a systematic review of the scientific literature on trustworthy and ethical Artificial Intelligence (AI) and Education (AI&ED), including both AI *applied* in education to support teaching and learning (AIED), as well as education *about* AI (AI literacy). Key interest is the identification of global trends with a special focus on unbalanced disparities. Strictly following the standardised protocol and the underlying PRISMA approach, 324 records were identified and selected according to the pre-defined protocol for the systematic review. Finally, 62 articles were included in the quantitative and qualitative analysis in response to four research questions: Which (i) journals, (ii) disciplines, and (iii) regions are leading scientific debates and sustainable developments in education and trustworthy/ethical AI, and (iv) what are the past trends? The articles revealed an unbalanced distribution across the various dimensions, together with an exponential growth over recent years. Building upon our analysis, we argue for an increase in interdisciplinary research that shifts the focus from the currently dominant technological focus towards a more human-centered (educational and societal) focus. Only through such a development AI can contribute effectively to the UN Sustainable Development Goal no. 4 of a world with equitable and universal access to quality education. The results of our systematic review provide the basis to address and facilitate equality in the future AI&ED progress across regions worldwide.

Keywords: Trustworthy and ethical AI, AI&ED, Web of Science articles, Systematic literature review, Informatics and information technologies, Education and learning sciences, Sustainable digital transformations.

1 Introduction

The concept of Artificial Intelligence (AI) has been controversial since the term was first coined [4, 14, 18]. Nonetheless, AI has been introduced in many disciplines, including – for around fifty years – in education [2, 13, 15, 20, 23]. However, it remains the case that AI in education was mostly researched by computer scientists rather than educators in the beginning [33].

2 Background

Educational systems and societies worldwide are increasingly challenged by rapid changes facilitated and caused by globalization, connectivity and new (social) media. In response, Open Education and Open Educational Resources (OER) have been promoted by the United Nations Educational, Scientific and Cultural Organization [29] to help achieve the United Nations' Sustainable Development Goal 4 (equitable and inclusive quality education for all) [25, 29]; and the value of OER was demonstrated during the COVID-19 lockdowns [26, 28, 30]. Meanwhile, it has also been suggested that recent technologies, such as AI, have potential for enhancing education, but they also bring various challenges [5, 6, 9, 13]. Several international agencies have discussed the potential of AI for future sustainable education [8, 19].

"AI in education for sustainable society" is the theme of the AIED 2023 Conference while the AIED 2024 Conference focuses "AI in education for a world in transition". Both themes call for societal considerations and objectives of future AI in education applications. However, to achieve a sustainable society, it is not only necessary to facilitate education *through* AI (the application of AI in education) but also to foster and improve education *about* AI (the teaching of AI in education) [10, 12, 13]. We need students and citizens who have digital competences, including what might be called 'AI Literacy', to understand, support and realize a sustainable society [11, 26, 28, 31]. Thus, we address both directions in our paper: AI applied in education (AIED) and AI taught in education (AI literacy) what we call "AI and Education (AI&ED)". We specifically focus on trustworthy and ethical AI&ED.

Already more than forty years ago, the research on using AI in education (AIED) began mainly focusing school and higher education [3, 13, 16, 27]. There are some systematic literature reviews on the current AIED research providing a first overview [5, 7, 17, 24, 32].

In the global AI&ED research and development community, ethical discussions were launched early but they were not gaining attention and continuation [27]. The identification of the necessity of ethical AI&ED and the development of community-driven proposals and frameworks for ethical AI&ED took twenty years [1, 2, 6, 11, 12].

To the best of our knowledge, the relation between trustworthy and ethical AI and education has not yet been systematically analyzed. This systematic literature review (SLR) aims to fill this gap for a topic that is likely to gain more importance in the near future. In addition, the results might inform future research and be used as a framework to differentiate and classify theoretical concepts and practical approaches. Accordingly, this work aims to explore the latest scientific literature trends concerning the relationship between trustworthy and ethical AI and education. To that end, we set out to answer the following four research questions:

RQ1: Which journals are leading scientific debates and sustainable developments in education and trustworthy/ethical AI?

RQ2: Which disciplines are leading scientific debates and ...?

RQ3: Which geographical regions are leading scientific debates and ...?

RQ4: Which trends are leading scientific debates and sustainable developments in education and trustworthy/ethical AI?

3 Methodology

The systematic review strictly followed the standardized protocol for systematic literature reviews on AI&ED [27] and the underlying PRISMA statement and its procedures [21, 22], which involves four phases for the selection of articles: (i) *Identification*, (ii) *Screening*, (iii) *Eligibility*, and (iv) *Included*. To ensure reliability, the four phases of the PRISMA process were undertaken by two reviewers (first two co-authors of this paper), each of whom have research experience in AI&ED and educational technology.

For the *Identification* phase, the reviewers reviewed in parallel the titles and abstracts of the records collected from the database Web of Science (Clarivate Analytics) using the search term: "TS = (("artificial intelligence") AND ((trust*) OR (ethic*)) AND (education))". The reviewers agreed on all records but one, and reached consensus on this final record after a discussion.

For the full *Screening* phase, all the titles and abstracts of the records generated by the first phase were reviewed, using the exclusion and inclusion criteria from [27].

For the *Eligibility* phase, the reviewers then reviewed in parallel the full text of the records remaining after the *Screening* phase, following the criteria defined by [27].

In the *Included* phase, all selected records were analysed related to the research questions.

4 Analysis and discussion

In this section, we present the results of the systematic review and their analysis in terms of the four research questions (RQs).

The *Identification* phase generated a list of 324 records (Table 1). In the *Screening* phase, 43 records were removed based on the analysis of their titles and abstracts and the criteria. In the *Eligibility* phase, 219 records were removed based on an analysis of the full texts and the criteria. This left a total of 62 scientific journal articles all of which were subject to quantitative and qualitative analyses.[1]

Table 1. Summary, with numerical results, of the four phases of the systematic review.

Phase	Screened records	Removed records
Identification	Records identified (n=324)	Duplicates removed (n=0)
Screening	Records for formal screening (title & abstract) (n=324)	Records removed by formal reasons (n=43)
Eligibility	Records for content-related screening (full text) (n=281)	Records removed by content-related reasons (n=219)
Included	Papers included for review analysis (n=62)	

[1] The selected 62 articles will be published with a DOI under an open and free license.

RQ1: Which journals are leading scientific debates and sustainable developments in education and trustworthy/ethical AI?

The 62 articles selected and reviewed for this SLR were published in 46 different journals, published by 23 publishing houses. The 'International Journal of the Artificial Intelligence in Education' (IJAIED) published most of the articles (8 in total), followed by 'Education and Information Technologies' (EIT) (4 publications) and the 'Journal of Research on Technology in Education' (JRTE) (3 publications). 39 out of the 46 journals published only one article. Based on the Web of Science statistics, the articles published in IJAIED have been cited on average 5 times (SD = 7.5), with 180 days usage count average of 11 (SD = 8). For the articles published in EIT the average 180 days usage count was 14 (SD = 8) but they were cited less (2 citations in total). For the articles published in JRTE the average 180 days usage count was 9 (SD = 10) while no citations were recorded. In terms of citations, the most cited article [32] was published in the 'International Journal of Educational Technology in Higher Education' (197 citations) followed by the article by [12] in the 'International Journal of the Artificial Intelligence in Education' with 23 citations.

RQ2: Which disciplines are leading scientific debates and ...?

To answer RQ2, we analyzed the articles included in this SLR in terms of the discipline of the publishing journal. Mainly, the articles were published in journals focusing on education, medicine and health care, technology and information systems, and business and social sciences. Some journals also had an interdisciplinary focus, such as 'Artificial Intelligence and Society' or 'Technology and Education'. Most articles were published in interdisciplinary journals with an education and technology focus (17 articles), followed by those with a technological focus (15 articles), an education focus (9 articles) and journals with strong medical and healthcare direction (8 articles).

RQ3: Which geographical regions are leading scientific debates and ...?

The regional distribution of the selected 62 articles is not balanced. The first authors of the articles are affiliated to 27 countries in total, although only 9 countries (USA: 13, UK: 7, China: 6, Australia & South Korea: 4, Finland & Spain: 3, Canada & Germany: 2) were represented by two or more papers. The selected 62 articles are also not equally spread across the globe, as is often seen due to different conditions and opportunities in relation to development and resources. Only five countries with ten articles in total belong to the so-called Global South, six of which are from China. Geographically, only two countries with a total of five articles belong to the southern hemisphere (Australia with four articles and South Africa with one article).

RQ4: Which trends are leading scientific debates and sustainable developments in education and trustworthy/ethical AI?

There are several interesting trends that we can derive from the quantitative and qualitative analysis of the 62 selected articles. Most obviously, there is the large growth of relevant articles during the most recent three years (Fig. 1.). Even though there was no time limit, the first publication only appeared in the year 1999 with a second one not following until 18 years later. Apart from the first article, all the other articles were published during the last five years with an exponential increase during the most recent three years. In short, it is reasonable to suggest that the discussion on trustworthy and ethical AI and education really only started three years ago.

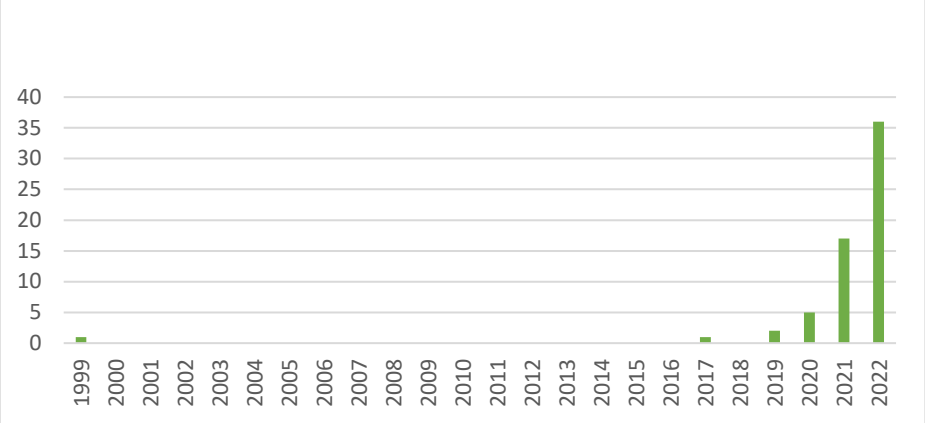

Fig. 1. Publication years of selected 62 articles of AI&ED systematic review.

Furthermore, trends can be identified in the analysis of the different article types. The vast majority of the selected 62 articles are discussions of theories, argumentations and literature (31 articles = 50.0 %), followed by ten survey analysis studies (16.1 %), seven systematic reviews (11.3 %), five mixed methods studies (8.0 %), four interview analysis studies (6.5 %), three thematic analysis studies (4.8 %), one data analysis study and one position paper of an association (1.6 % each).

5 Discussion

In this section, we discuss the results and analysis of the selected 62 articles as presented in the sections before. We structure again our discussion along the leading four research questions (RQs).

RQ1: Which journals are leading scientific debates and sustainable developments in education and trustworthy/ethical AI?
Given the basis of 62 analysed articles, the huge number of different journals (46) is indicating a pre-mature research field without any leading publication channels. Only 3 journals (7 %) count more than 2 articles while the vast majority (39 of 46 = 85 %) has published only one article from the selected 62 items. That underlines that the topic trustworthy and ethical AI&ED is not yet assigned to specific journals which have also not discovered it as potential key focus for distinction from other journals. It is striking that only one article achieved high visibility and citations while even the article on the second position presents a huge distance related citation (197 against 23 citations).

RQ2: Which disciplines are leading scientific debates and sustainable developments in education and trustworthy/ethical AI?
Not surprisingly, most of the selected 62 articles (41 = 66 %) are belonging to disciplines summarized in education and technology and all other disciplines are represented only one time except medicine and healthcare (8 = 13 %) and (social)

sciences (10 = 16 %). This huge representation of medicine and healthcare is representing the big efforts to implement and use AI&ED in the health system and education what has to be combined with ethical reflections and statements as mandatory requirement of the discipline. On the other hand, the even higher representation of (social) sciences is remarkable as this discipline is not known for many AI&ED implementations and uses: We can argue that the discussion of trustworthy and ethical concerns is very common for (social) sciences

RQ3: Which geographical regions are leading scientific debates and sustainable developments in education and trustworthy/ethical AI?
The distribution in relation to country affiliations of the first authors demonstrates a clear unbalance between developed (52 = 84 %) against the rest of the world (10 = 16 %) as well as between northern (60 = 97 %) and southern (2 = 3 %) countries. This suggests again that the development, research and implementation of innovative technologies such as AI are mainly (presumably for socio-economic reasons) driven by countries in the Global North in deep contrast to the Global South. That is independent from the ongoing debate whether China should be assigned to the Global South what it questionable given the economic power and progress of this special country due to its largest population worldwide. Overall, all these conditions are challenging the UN Sustainable Development Goals (SDG) and in particular the achievement of SDG no. 4 which demands equitable and universal access to quality education worldwide.

RQ4: Which trends are leading scientific debates and sustainable developments in education and trustworthy/ethical AI?
The rise of the research on trustworthy and ethical AI&ED started only three years ago as already mentioned in the analysis. However, the almost incredible increase by 2000 % within three years is indicating a dramatic disruption and shift in the research focus. One reason could be seen in the similar changes in business investments, funding projects, opened job opportunities (both in enterprises and academia) and established professorships and research positions.
The complete absence of pedagogical, design-based as well as empirical, evidence-based studies on experiments from the practice is notable and underlines the starting point of a scientific debate that is still lacking practical developments and implementations for research.

Overall discussion
This SLR has revealed that the scientific debate on trustworthy and ethical AI and education is just in its infancy. This is especially noteworthy given the long history of AI&ED research, and the almost ten years of research and publications centered on the ethics of AI in general. Accordingly, we argue that research into trustworthy and ethical AI&ED research requires more effort, in particular in terms of pedagogical, design-based as well as empirical, evidence-based studies of ethical AI in educational practice. In addition, this SLR has shown that the majority of relevant articles have been published in journals focused on technology-oriented education, notable as most of the authors are not from the discipline/field of education. Meanwhile, only one of the 62

reviewed articles addresses AI literacy, which is therefore another research gap that needs to be addressed. Thus, we encourage AI&ED researchers worldwide to undertake studies to test and evaluate practices of ethical AI in education and to explore what a robust AI literacy might mean (addressing both the technological and human dimensions of AI [13]), to help strengthen UN SDG 4.

Limitations

We highlight two limitations of this SLR. First, we collected records only from one database (Web of Science) which revealed only 62 articles. Had we used additional databases, more articles might have been identified. Second, the selection processes were subjective, based on the personal judgements of two researchers (although, to minimize this, as noted earlier we analyzed and compared outputs at two stages of the process, and discussed and resolved any cases that were not clear-cut).

6 Conclusion

This SLR provides the first overview of the published scientific literature on the relationship between trustworthy and ethical AI and education. It addresses four research questions focused on (i) the most common journals, (ii) the most common disciplines, and (iii) the most common regions leading scientific debates and sustainable developments on trustworthy/ethical AI and education, and (iv) past trends. The key finding is that ethical AI&ED analyses are only just starting to appear, which is why we call for more efforts in this increasingly important area.

We argue that research into ethical AI&ED can and should accompany more general AI&ED research and developments, and should involve testing and evaluation of ethical AI&ED practices. In particular, there is a need for pedagogical, design-based as well as empirical, evidence-based studies of ethical AI in educational practice and of AI literacy. In this way, this community can better contribute to sustainable learning practices, to foster and strengthen sustainable equitable and inclusive quality education for all (UN SDG 4) and our future societies and citizens.

References

1. Akgun, S., & Greenhow, C. (2021). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics, 2*, 431–440. https://doi.org/10.1007/s43681-021-00096-7
2. Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics, 1*(1), 61–65. https://doi.org/10.1007/s43681-020-00002-7
3. Bozkurt, A., Xiao, J., Lambert, S., Pazurek, A., Crompton, H., Koseoglu, S., ..., & Jandrić, P. (2023). Speculative Futures on ChatGPT and Generative Artificial Intelligence (AI): A Collective Reflection from the Educational Landscape. *Asian Journal of Distance Education, 18*(1), 53-130. https://doi.org/10.5281/zenodo.7636568
4. Chaka, C. (2023). Fourth industrial revolution—a review of applications, prospects, and challenges for Artificial Intelligence, robotics and blockchain in higher education. *Research and Practice in Technology Enhanced Learning, 18*(2), 002. https://doi.org/10.58459/rptel.2023.18002
5. Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510
6. Chounta, I.-A., Bardone, E., Raudsep, A., & Pedaste, M. (2022). Exploring teachers' perceptions of Artificial Intelligence as a tool to support their practice in Estonian K-12 education. *International Journal of Artificial Intelligence in Education, 32*(3), 725-755. https://www.doi.org/10.1007/s40593-021-00243-5
7. Crompton, H., Jones, M. V., & Burke, D. (2022). Affordances and challenges of artificial intelligence in K-12 education: a systematic review. *Journal of Research on Technology in Education*. https://doi.org/10.1080/15391523.2022.2121344

8. European Commission (2022). *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*, https://data.europa.eu/doi/10.2766/153756
9. European Parliament (2021). Report on artificial intelligence in education, culture and the audiovisual sector (2020/2017(INI)). https://www.europarl.europa.eu/doceo/document/A-9-2021-0127_EN.html
10. Holmes, W. (2023). *The Unintended Consequences of Artificial Intelligence and Education.* Education International Research. https://www.ei-ie.org/en/item/28115:the-unintended-consequences-of-artificial-intelligence-and-education
11. Holmes, W., Persson, J., Chounta, I.-A., Wasson, B., & Dimitrova., V. (2022a). *Artificial Intelligence and Education. A critical view through the lens of human rights, democracy and the rule of law*, https://rm.coe.int/artificial-intelligence-and-education-a-critical-view-through-the-lens/1680a886bd
12. Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., …, & Koedinger, K. R. (2022b). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education 32*(3), 504-526. https://www.doi.org/10.1007/s40593-021-00239-1
13. Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education, 57*, 542– 570. https://doi.org/10.1111/ejed.12533
14. Huang, R., Tlili, A., Xu, L., Chen, Y., Zheng, L., Saleh Metwally, A. H., ..., & Bonk, C. J. (2023). Educational futures of intelligent synergies between humans, digital twins, avatars, and robots - the iSTAR framework. *Journal of Applied Learning & Teaching, 6*(2), 1-16. https://doi.org/10.37074/jalt.2023.6.2.33
15. Ifelebuegu, A. O., Kulume, P., & Cherekut, P. (2023). Chatbots and AI in Education (AIEd) tools: The good, the bad, and the ugly. *Journal of Applied Learning and Teaching, 6*(2). https://doi.org/10.37074/jalt.2023.6.2.29
16. Kent, C., & du Boulay, B. (2022). *AI for Learning*. CRC Press, Boca Raton, FL. https://doi.org/10.1201/9781003194545
17. Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education, 30*, 121-204. https://doi.org/10.1007/s40593-019-00186-y
18. McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955). *A proposal for Dartmouth Summer Research Project on Artificial Intelligence*. https://www-formal.stanford.edu/jmc/history/dartmouth.pdf
19. Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000386693
20. Mills, A., Bali, M., & Eaton, L. (2023). How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process. *Journal of Applied Learning and Teaching, 6*(1). https://doi.org/10.37074/jalt.2023.6.1.34
21. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097
22. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., …, & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews, 10*, 89. https://doi.org/10.1186/s13643-021-01626-4
23. Pinkwart, N. (2016). Another 25 years of AIED? Challenges and opportunities for intelligent educational technologies of the future. *International Journal of Artificial Intelligence in Education, 26*, 771-83. https://doi.org/10.1007/s40593-016-0099-7
24. Sanusi, I.T., Oyelere, S.S., Vartiainen, H., Suhonen, J., & Tukiainen, M. (2022). A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies*. https://doi.org/10.1007/s10639-022-11416-7
25. Stracke, C. M. (2019). Quality Frameworks and Learning Design for Open Education. *The International Review of Research in Open and Distributed Learning, 20*(2), 180-203. https://doi.org/10.19173/irrodl.v20i2.4213
26. Stracke, C. M., Burgos, D., Santos-Hermosa, G., Bozkurt, A., Sharma, R. C., Swiatek, C., ..., & Truong, V. (2022a). Responding to the initial challenge of COVID-19 pandemic: Analysis of international responses and impact in school and higher education. *Sustainability, 14*(3), 1876. https://doi.org/10.3390/su14031876
27. Stracke, C. M., Chounta, I.-A., Holmes, W., Tlili, A., & Bozkurt, A. (2023). A standardised PRISMA-based protocol for systematic reviews of the scientific literature on Artificial Intelligence and education (AI&ED). *Journal of Applied Learning and Teaching, 6* (2). https://doi.org/10.37074/jalt.2023.6.2.38
28. Stracke, C. M., Sharma, R. C., Bozkurt, A., Burgos, D., Swiatek, C., Inamorato dos Santos, A., ..., & Truong, V. (2022b). Impact of COVID-19 on formal education: An international review on practices and potentials of Open Education at a distance. *The International Review of Research in Open and Distributed Learning, 23*(4), 1-18. https://doi.org/10.19173/irrodl.v23i4.6120
29. United Nations Educational, Scientific and Cultural Organization (2019, November 25). UNESCO Recommendation on OER. http://portal.unesco.org/en/ev.php-URL_ID=49556&URL_DO=DO_TOPIC&URL_SECTION=201.html
30. UNESCO, UNICEF, The World Bank, & OECD. (2021, June). *What's next? Lessons on education recovery: Findings from a survey of ministries of education amid the COVID-19 pandemic*. http://covid19.uis.unesco.org/wp-content/uploads/sites/11/2021/07/National-Education-Responses-to-COVID-19-Report2_v3.pdf
31. Vuorikari, R., & Holmes, W. (2022). DigComp 2.2. Annex 2. Citizens Interacting with AI Systems. In R. Vuorikari, S. Kluzer, & Y. Punie, *DigComp 2.2, The Digital Competence framework for citizens: With new examples of knowledge, skills and attitudes* (pp. 72–82). https://data.europa.eu/doi/10.2760/115376
32. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0