Guidelines for the use of artificial intelligence in university contexts

Prepared by Juan David Gutiérrez (juagutie@uniandes.edu.co), associate professor at the School of Government, Universidad de los Andes. Version 5.0 (15/08/2023).[1]

1. Introduction

- **We use artificial intelligence tools daily.** Consciously or unconsciously, practically all of us use some kind of artificial intelligence (AI) in our daily lives, in particular systems based on language models. For example, we use language models when we are writing a text message or an email and the "autocomplete" tool suggests how to finish writing the sentence. We also use language models to support us in carrying out academic activities, such as automated translation systems or programs that detect grammatical errors, among others. In addition, nowadays, there are web platforms and mobile applications that provide access to AI tools to generate texts and audiovisual content.

- **Some uses of AI tools can support learning, research, teaching, and other professional activities.** There are appropriate uses of such technologies, which can not only be useful to support learning, teaching, and research processes at the university but also at a professional level. For example, different AI tools nowadays contribute to activities at all stages of public policy processes, and, in Colombia, more than 50 public sector entities have adopted automated decision systems to support the performance of their functions.

- **But certain uses of AI are risky, so their use must be informed, transparent, ethical, and responsible.** AI tools are not suitable for every type of activity and certain types of use can be counterproductive to the pedagogical process. Some uses of AI can generate risks for users and third parties. Furthermore, as explained in section 4 of these guidelines, it is essential to be aware of the ethical, environmental, and human rights implications associated with the use of these tools. It is precisely because of these risks that some universities and professors have published guidelines, recommendations, principles, policies, and directives on the use of AI in academic contexts (see list in section 5 of these guidelines).

- **We should not assume that all our students are "digital natives" or that they want to adopt new technologies.** People's attitudes towards technologies are diverse and complex, it should not be assumed that all our students are eager to adopt new technologies. For example, in the surveys I have carried out in workshops with students on the use of AI at university, I have found that these systems generate very diverse feelings in them: fear, intrigue, excitement, uncertainty, curiosity, distrust, euphoria, worry, amazement, expectation, anguish, tranquility, confusion, fascination, caution, questioning, rejection, uneasiness and comfort. Moreover, people's capacity to adopt new technologies is heterogeneous and depends on different economic, social and cultural factors, among others. All the above implies that those of us who teach should implement differentiated teaching strategies according to the needs of our students and that we should not force the use of new AI systems.

[1] I wish to thank my research assistant Sarah Muñoz-Cadena for helping me with the translation of the guidelines. The Spanish version is available here. I also thank Professor A. Mills for proof-reading an earlier version of the guidelines.

- **Guidance for the use of AI inside and outside the university classroom.** These guidelines provide basic guidance for university professors and students on how to use AI, particularly large-scale language models, in an informed, transparent, ethical, and responsible manner. The challenges of using AI in education are not limited to aspects associated with learning processes and academic research. For example, the adoption of AI to (semi)automate personnel hiring and evaluation processes or for the surveillance of educational facilities also has legal and ethical implications. But these issues are beyond the scope of these guidelines since the rules and principles outlined here focus exclusively on the use of AI for educational and academic purposes inside and outside the university classroom.

- **Definitions of three key terms.** The following are working definitions of three technological terms used in these guidelines: **(A)** Artificial Intelligence (AI) System: "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions that influence real or virtual environments. AI systems are designed to operate with varying levels of autonomy". **(B)** Language model: algorithms that process, analyze and produce natural language in written or spoken form. **(C)** Large Language Model (LLM): algorithms trained on a vast amount of information whose main function is to predict the most likely string of words given the context that precedes or follows it.

2. Objectives of these guidelines

- **Contribute to digital literacy.** Contribute to learning basic knowledge about the use of AI in university contexts – with emphasis on LLMs. This objective includes teaching about its benefits and risks, its implications for society, and the development of critical digital competencies.

- **Promotion of responsible use of AI.** Promote the responsible and ethical use of these technologies in learning processes and for future use in professional life.

- **Prevent the violation of academic integrity.** Prevent situations in which students consciously or unconsciously incur in academic dishonesty.

3. Rules for the use of AI in and out of the classroom

- **General rule.** The use of AI as a support tool to carry out different learning activities is allowed. The parameters for the use of these tools described below distinguish between "low risk" and "high risk" tools, depending on the risks that the respective AI generates for the pedagogical process, for users and for third parties.
- **Rules for the use of low-risk AI**. Students may freely use low-risk AI tools, i.e., those that allow to correct or review student-generated content or those that allow to collect and process data. For example, grammar correction tools, translation tools, audio-to-text transcription tools, and Internet information search tools, among others. Where such tools enable students to do meaningful work (e.g., translation), they must indicate their use in the appropriate section (e.g., methodology section). If students have doubts about how to classify the risk of using a specific AI, I suggest contacting their teacher.

- **Rules for the use of high-risk AI.** AI tools that are used to generate synthetic text are high-risk. That is, LLMs that can be accessed through chatbots, such as *ChatGPT*, *Bard*, *Bing*, and *Claude*, etc. Students can include AI-generated content in their individual and collective deliverables when four requirements are met:

  i. **Informed use.** Prior to using the tool, students must investigate who or what company developed the tool, how it was developed, how it works, what functions it can perform, what its terms of use are, what the company does with the information it collects from the user, and what limitations and/or risks it may generate for the user and/or third parties.

  ii. **Transparent use.** Students must explain in detail and expressly which tool were used and how the tools were used in the methodology section or in the section that they deem pertinent. This is a requirement currently requested by scientific journals and publishing houses, such as Nature and Cambridge University Press, to whom manuscripts are submitted for peer review.

  iii. **Ethical use.** In relation to the use of LLMs, students must distinguish what was written or produced directly by them and what was generated by an AI tool. On this point, general citation rules apply, e.g., use quotation marks if textual paragraphs are included. For citing synthetic text produced by LLMs, students may consult the rules suggested by the APA style team. Violations of this rule, particularly with respect to the second requirement, will be dealt with under an approach like that which applies in cases of plagiarism. In addition, ChatGPT is not a person, so it cannot be considered a co-author. Some scientific journals have already updated their publication policies to clarify that language models such as ChatGPT do not satisfy their "authorship" attribution criteria. On the other hand, the ethical use of AI systems – in general, not just LLMs – should be informed by four principles of bioethics: beneficence (achieving benefits), respect for the autonomy of persons (including obtaining consent if third parties are involved), justice (especially non-discrimination), and non-maleficence (not intentionally or recklessly causing harm to others).

  iv. **Responsible use.** The use of AI tools should be limited to early stages of research, to inspire or suggest directions, not to produce content that will later be included in students' deliverables. In any case, if students choose to transcribe texts produced by an LLM they must prove that the content was rigorously checked with reliable sources, since these systems tend to offer inaccurate, erroneous and false information. Responsible use also implies avoiding that the use of the system does not cause damage to the user or third parties. For example, it is not recommended to include personal data or confidential information about the students or others in chatbot queries because once such information is entered, the user loses control of who can access it and there is a risk that it will be known by unauthorized third parties.

- **Plausible uses of LLMs:**

  i. **Explore.** To explore new issues and seek inspiration, i.e., use the chatbot in the earliest stages of research.

  ii. **Automate basic tasks.** To support exercises that involve recognition and/or reproduction of patterns in texts, for example: translating, summarizing, modifying the tone and style of the text – for example, to make the text clearer or more accessible to certain audiences –, correcting spelling and grammar, audio-text conversion, programming, etc. Students should no rely on automated tasks, the LLM's output must be carefully reviewed.

  iii. **Trivial writing.** To support unimportant writing tasks. To decide whether something is trivial or not, think about the following heuristic: "what would happen if my interlocutor found out that I used an LLM to write the message?".

  iv. **Sparring.** Testing one's own arguments or ideas through interaction with the chatbot, for example, by asking the system to respond with criticisms or counterarguments to the student's statements.

- **Not recommended uses of LLMs:**

  i. **Use the chatbot as the main or only source to search for factual or technical information.** I do not recommend it, LLMs are not reliable search tools. Besides, Internet search engines are more efficient and reliable for that task. In any case, since searching chatbots can be perceived as an easier and faster way to obtain information, I recommend that the information obtained through chatbots should always be carefully checked with reliable sources.

  ii. **Use of synthetic text for substantive issues.** I do not recommend it because of the tendency of LLMs to produce answers with inaccurate, erroneous and false information. In any case, if students do so, each point should be checked against other reliable sources.

  iii. **Doing mathematical calculations.** LLMs are not trained to perform mathematical calculations and thus often fail to answer simple arithmetic operations.

- **Accompaniment of the teacher and development of digital competencies.** The teacher will dedicate class time to explain what AI is, what are LLMs that power tools like ChatGPT, what opportunities and risks do these systems generate for academic and professional work, and what are the ethical, environmental, and human rights implications associated with the use of these tools. The professor will be available to clarify the scope of these guidelines, to discuss and co-create them, and to resolve specific questions about the use of AI. In addition, the teacher will seek to carry out activities that contribute to the digital literacy of students, particularly for the development and strengthening of digital competencies such as the ability to critically evaluate the benefits and risks of new technologies.

- **Supervision**. The teacher will use different sources to identify potential situations of academic dishonesty using AI, but he or she will not rely on the tools available to detect synthetic text since none of these tools are currently reliable. In any case to avoid risks associated with academic integrity, the main strategy the teacher will use will be pedagogical not punitive.

4. Why the informed, transparent, ethical, and responsible use of AI in academia is necessary?

- **The use of AI tools must be informed, transparent, ethical and responsible for – at least – four types of reasons: (1)** because LLMs are not always reliable; **(2)** because there are risks that certain uses may negatively affect learning processes; **(3)** because of the risk of users treating AI as if it were a human being; and **(4)** because the use of the tools has ethical and human rights implications due to the way they were developed and/or because some tools may replicate or amplify social issues such as discrimination. Each of the four reasons is explained in detail below:

- **First**, because their answers are NOT always reliable even though, for example, an AI generator produces text that looks convincing. LLM-based systems such as ChatGPT do not perform with the accuracy of other tools used in learning environments such as calculators. In fact, ChatGPT tends to include false or fanciful information in its responses. Microsoft's built-in chatbot for its Internet browser, Bing, and Google's chatbot, Bard, have also presented the same types of problems. These systems do not distinguish true from false. Why does this happen? LLMs spin words from probabilistic inferences from the data they were trained on, but they do not have the ability to understand what they produce, nor do they associate meanings to the words they output (they are "stochastic parrots").

- **Second**, because the use of text-generating tools may discourage students' motivation to write and think on their own. It is worth reiterating that the learning activities in this subject seek to develop their cognitive skills and that these guidelines seek to prevent some AI tools from becoming automated plagiarism mechanisms.

- **Thirdly**, because of the risk that users, consciously or unconsciously, treat AI behavior as if it were human (Eliza Effect). For example, LLM-based tools do not "understand" the texts they produce; they mimic language patterns by synthesizing large volumes of data based on which they generate word sequences. This problem of anthropomorphizing machines is exacerbated by the fact that some chatbots have produced violent or harassing responses that could lead people to engage in harmful behavior.

- **Fourth**, because of the ethical and human rights implications associated with the use of certain AI systems given that: some tools tend to reproduce or amplify derogatory and discriminatory stereotypes associated with gender, race, ethnicity or disability; technologies could have been developed from massive copyright infringement; new forms of colonialism through the non-consensual extraction of information from historically marginalized communities; some tools would have been developed in contexts of labor exploitation; the development and operation of such systems generates a considerable

carbon footprint; and, the potential violation of privacy and personal data protection rights of those who use them. For more information on the harms associated with the use of Generative AI, I recommend consulting the report "Ghost in the machine: Addressing the consumer harms of generative AI" (2023) by the Norwegian Consumer Council and the report "Generating Harms: Generative AI's Impact & Paths Forward" by the Electronic Privacy Information Center.

5. Other resources for teachers

- **Other guidelines on the use of AI at the university.** I recommend consulting guidelines, policies, and principles of other universities, professors, and other organizations, such as: University of Washington, Hertie School, University of Tartu, Yale University, University of Helsinki, Technical University of Munich, University College London, UNESCO, UK Department of Education, Boston University, Warwick University, Universidad de Buenos Aires (IALAB), Universidad del Desarrollo, Russell Group, and Lance Eaton's list.

- **Introductory literature on LLMs. (A)** For an introduction on how LLMs were developed and how they work (in particular, GPT-3), I recommend this text by Timothy B. Lee and Sean Trott. **(B)** For an explanation on how LLMs work and what are their main risks, I recommend the text "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" by Emily M. Bender, Timnit Gebru, Angelina McMillan-Major and Shmargaret Shmitchell.

- **Teaching resources. (A)** For those interested in the challenges that LLMs create for teaching writing in universities, I suggest consulting the paper entitled "Adapting College Writing for the Age of Large Language Models such as ChatGPT: Some Next Steps for Educators" by Anna Mills and Lauren M. E. Goodlad. **(B)** I also suggest the materials contained in "AI Text Generators and Teaching Writing: Starting Points for Inquiry" by Anna Mills. **(C)** The article by Anna Mills, Maha Bali and Lance Eaton entitled "How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process" documents and reflects on how professors from different universities around the world have adopted new pedagogical practices to respond to the challenges generated by Generative AI. **(D)** MIT offers for free its DAILy Curriculum program that contains materials to teach basic AI skills to teenagers.

- **Critical approach to the use of AI in academic settings. (A)** I recommend consulting Maha Bali's list of resources on developing critical literacy in AI. **(B)** I also suggest reading a reflection by Mohammad Hosseini, Lex Bouter, and Kristi Holmes in which they argue that it is not desirable to adopt AI tools in education without first reflecting on their effects and biases and without having adopted measures to mitigate their risks. **(C)** Naomi S. Baron's article "How ChatGPT robs students of motivation to write and think for themselves" discusses how the use of LLMs can negatively affect students' creative processes. **(D)** Dimitrinka Atanasova's blog post reflects on how LLMs can reduce barriers for students and researchers to write in English (when this is not their first language), but that their effects on cultural inclusivity in higher education are more ambiguous. **(E)** Finally, I recommend the Kathryn Conrad's text "A Blueprint for an AI Bill of Rights for

Education" that proposes rights for teachers and students that can inspire policy formulation and protective measures.

6. Guidelines open to change

- AI is a rapidly changing set of tools, which is why these guidelines will remain open to future evaluations, modifications, and revisions.
- Throughout the academic semester, the professor will open spaces to discuss these guidelines with the students and, if necessary, modifications may be introduced through co-creation exercises.
- I am grateful to my colleagues and others who provided comments and criticisms on initial versions of this document and also to those who have participated in the talks and workshops I have given on the use of large-scale language models in university contexts.

7. Changes introduced in this version of the guideline

- I introduced changes of form and substance with respect to the previous version of this instrument (v. 4.3), which I summarize below.
- **Changes of form: (A)** The previous version was called "Guidelines for the Use of Artificial Intelligence in University Courses", the new title is broader since the instrument is useful for academic areas that go beyond classroom activities (e.g., research activities). **(B)** I introduced new sources that can be consulted in the form of hyperlinks throughout the text.
- **Substantive changes: (A)** In point 1 of the new version of the guidelines I included a glossary of key terms and in point 3 I introduced a new sub-section with plausible and non-recommended uses of LLMs by students. **(B)** In addition, in this version I included a new section of resources for teachers that includes examples from other guidelines, basic literature on LLMs, resources for teaching, and critical literature on the use of AI in educational contexts. **(C)** On the other hand, the previous version of this instrument indicated that the teacher would use synthetic text detection systems, but I have removed those allusions in the present version because the currently available systems produce a high level of false positives and negatives. For this reason, for example, the Teaching Center at the University of Pittsburgh issued a press release stating that it did not endorse or promote any of the tools currently available.

8. Use of these guidelines in your classes

- These guidelines are published under a Creative Commons Attribution 4.0 license, so if you are an educator interested in formulating your own policies or guidelines, you can share and adapt these guidelines as long as you make the corresponding attribution and indicate if you have made any changes.
- If you would like to send me your comments, criticisms, and suggestions, please write to me at juagutie@uniandes.edu.co .