

Don't Be Sorry, Just Declare It: Pedagogical Principles for the Ethical Use of ChatGPT, Master Bullshit Artist of Our Time

Benito Cao 

Department of Politics and International Relations, The University of Adelaide, Australia.

How to cite: Cao, B. (2025). Don't Be Sorry, Just Declare It: Pedagogical Principles for the Ethical Use of ChatGPT, Master Bullshit Artist of Our Time. In: 11th International Conference on Higher Education Advances (HEAd'25). Valencia, 17-20 June 2025. <https://doi.org/10.4995/HEAd25.2025.20092>

Abstract

This work aims to respond to the pedagogical challenges posed by ChatGPT. The first part examines ChatGPT from a critical perspective, deploying the philosophical work of Harry Frankfurt "On Bullshit" (2005) to conclude that ChatGPT is the master "bullshit artist" of our time. The second part illustrates the integration of four principles to promote its ethical use: caution, trust, relevance, and transparency. These principles are informed by industry standards and, in essence, come to reflect the advice provided by Australian Customs and Biosecurity to people who arrive in Australia: "Don't Be Sorry, Just Declare It." The third part provides evidence of the implementation of this approach in courses with a combined enrolment of 214 students over two successive years. The evidence indicates that students respond well to this pedagogical initiative and reveals that this approach can minimise the potential cases of academic misconduct.

Keywords: Artificial Intelligence; ChatGPT; Pedagogy; Academic Integrity; Ethics

1. Introduction

Artificial intelligence (AI) has become an integral part of our lives and a powerful and ubiquitous pedagogical tool. As generative AI technologies are increasingly employed in educational settings, it becomes imperative to ensure their responsible and ethical use. This paper is partly a provocation, but one aimed at enhancing our understanding of (and inspiring our critical engagement with) ChatGPT, the most popular manifestation of generative AI. The first part of the paper reflects on ChatGPT from a critical perspective, deploying the philosophical work of Harry Frankfurt *On Bullshit* (2005) to conclude that ChatGPT is the master bullshit artist of our time. The second part proposes the integration of four principles to promote the ethical use of ChatGPT: caution, trust, relevance, and transparency. The third and final part illustrates these principles in practice and provides evidence of their implementation in courses with a combined enrolment of 214 students over two successive years.

2. ChatGPT - The Bullshit Artist at Work

The public release of ChatGPT in November 2022 has dramatically reshaped the way people interact with AI by making advanced conversational intelligence widely accessible. The so-called ChatGPT revolution has accelerated automation, transformed the education sector, and sparked global discussion on the opportunities and challenges posed by AI (e.g. Bobula, 2024; Kaur & Shilpa, 2024; Kurban & Şahin, 2024). The main challenge comes from the generative dimension of these new AI technologies, which allows them to create content (e.g. text, images, music, and code) based on patterns learned from vast datasets. Models like ChatGPT, for instance, assemble words based on probabilities derived from training on large amounts of text, allowing them to produce human-like responses. This presents huge pedagogical challenges, particularly regarding assessment (e.g. Swiecki *et al*, 2022), that require, first and foremost, a proper understanding of the nature of generative AI tools like ChatGPT

Here, I am proposing that the operating nature of ChatGPT is that of a master bullshit artist. To articulate this point, let me share a personal story. In early 2023, confronted with the fact that students could make use of ChatGPT to produce their essays, and having decided I needed to do something about this, I realised that most students (and most of my colleagues) assumed that ChatGPT operates like a search engine, navigating the internet in search of information to answer their questions. Determined to understand the technology, I quickly learned that ChatGPT is not a search engine. ChatGPT is a large language model that produces text based on statistical probabilities. To explain and illustrate this to my students, I had the idea of prompting ChatGPT to provide a profile of myself, using my full name: Benito Cao Lopez.

Despite the fact that a detailed profile of myself is publicly available in the University of Adelaide website, and is the first content that comes up when one does a Google search of my full name, the profile generated by ChatGPT was riddled with errors and generalisations.

This is an annotated example of a profile created with a free version of ChatGPT in 2023:



Associate Professor Benito Cao Lopez is a highly accomplished and esteemed individual in the field of academia. Benito Cao Lopez was born on October 15, 1975, in Madrid, Spain [wrong date, year and location]. Benito grew up in a supportive and nurturing environment, where his parents encouraged his curiosity and provided him with the necessary resources to explore his interests [not even close: poor parents; broken family; grew up in an orphanage]. Benito completed his undergraduate studies at the Autonomous University of Madrid, where he obtained a Bachelor of Science degree in Physics [wrong university and degree]. Benito pursued a Ph.D. in Theoretical Physics at the prestigious University of Cambridge, United Kingdom [wrong degree and university]. He began his professional journey as a postdoctoral researcher at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland [never been there]. After his tenure at CERN, Benito joined The University of Adelaide as a faculty member in the Department of Physics [wrong department]. His research interests encompass various areas of theoretical physics, including quantum field theory, particle physics, and cosmology [ridiculously inaccurate].

This is a perfectly plausible profile of someone called Benito Cao Lopez, but it is not an accurate profile of this Benito Cao Lopez. Indeed, it is not an accurate profile of any Benito Cao Lopez out there. I tried several variations of this exercise and the only accurate content was that I was born in Spain and employed at the University of Adelaide. The rest were incorrect variations of my date and city of birth, my academic studies, and my areas of interest and expertise. I tried again recently and apparently I am now an expert in Applied Linguistics. These are all different profiles about the same person, and they are all ridiculously inaccurate, but they are always communicated in an authoritative voice, with the confidence of someone who knows me, who knows the Benito Cao Lopez currently employed at the University of Adelaide.

Going over such ridiculously inaccurate profiles, I caught myself repeatedly yelling: Bullshit! This reminded me of Harry Frankfurt's book *On Bullshit* (2005). I revisited the text and realised that ChatGPT was the perfect illustration of a bullshit artist at work. Frankfurt begins by making the point that: "One of the most salient features of our culture is that there is so much bullshit. Everyone knows this. Each of us contributes his share. But we tend to take the situation for granted" (2005: 1). He then proceeds to differentiate bullshitting from lying, in that unlike lying, bullshitting is unconcerned with the truth. He argues that bullshitting involves a kind of bluff and thus, like bluffing, it is more a matter of fakery than falsity. He states: "the essence of bullshit is not that it is *false* but that it is *phony*" (2005: 147). He concludes by noting that bullshit is unavoidable whenever someone talks without knowing what they are talking about.

This is exactly what ChatGPT does. ChatGPT generates text about anything and everything with no understanding of the content, but it does so with the confidence of an all-knowing expert. This makes ChatGPT the master bullshit artist of our time. The notion that ChatGPT is a bullshit generator has been articulated by other authors (e.g. Hicks *et al*, 2024). However, this alone does not capture the full epistemological impact of this technology. Its impact does not lie merely in its unprecedented capacity to generate bullshit (commonly understood as false or misleading content, and often referred to as hallucinations in the world of generative AI), but on the fact that it presents that content with a confident and authoritative voice. In other words, the key is not to see ChatGPT merely as a bullshit generator, but as a master "bullshit artist".

Note that this does not mean that ChatGPT cannot produce accurate content. It can, and it often does. Predictive text in our smartphones sometimes (perhaps even often) correctly identifies the next word we intend to write, but this is not because the algorithm knows what we intend to write. Predictive text is merely a probabilistic tool -a very simple language model, if you will. In a sense, ChatGPT is predictive text on steroids -a very sophisticated large language model. ChatGPT has no understanding of anything, and thus it cannot tell whether the content it processes and the content it produces is accurate or not. This is why pedagogical reliance on ChatGPT is fraught with danger, and this is something our students must understand so they take a careful and critical approach to this technology. The key advice I provide to students regarding ChatGPT (and generative AI in general) is simple: Caution – Bullshit Ahead.

3. The Urgent Need for Normative Principles

The status of ChatGPT as the quintessential bullshit artist underscores the urgent need for normative principles to address the challenges posed by its use in educational settings. Here, I am proposing the integration of four principles: caution, trust, relevance, and transparency. I am not suggesting these are the only relevant principles, or even the most important ones. I am merely proposing these principles as a starting point to address the pedagogical challenges posed by ChatGPT, particularly regarding the production of written assignments. Briefly, this is how I articulate each of these principles in the context of my teaching practice.

3.1. Caution

Students should be warned about the risks of using ChatGPT in their work and encouraged to cultivate a critical approach when engaging with the tool, particularly in relation to truth. As established earlier in this paper, ChatGPT has no regard for truth. It does not lie or tell the truth. It simply does not know the difference and does not care. Indeed, the technology is “designed to optimise the believability of an answer rather than its accuracy” (Messerli, 2023: 90). Thus, whilst generative AI tools can accelerate content creation, their use can undermine the learning process if students are not taught to be careful and critical. Educators must guide students in approaching generative AI with caution and using it responsibly, noting the importance of verifying AI-generated content and encouraging evaluative judgement (e.g. Bearman *et al*, 2024). In addition to containing its pedagogical risks, this can help students cultivate epistemic doubt and enhance their capacity for critical thinking (e.g. Davies *et al*, 2024).

3.2. Trust

Trust between students and teachers is essential for effective education (e.g. Curzon-Hobson, 2002; Bormann *et al*, 2021). Trust fosters an open learning environment where students feel confident seeking guidance, engaging in critical discussions, and complying with assessment guidelines. In a context in which “there is discernible concern among the higher education community that GenAI is causing an ‘erosion of trust’ between students and their teachers” (Luo, 2024: 2), teachers must foster a trusting learning environment when it comes to using ChatGPT. This can be done by engaging students during the development of guidelines about appropriate and inappropriate uses of generative AI tools, as well as by providing clear and justified reasons for those guidelines. Regulatory and punitive approaches still have a place in higher education, particularly in misconduct scenarios, but developing a trusting learning environment, particularly around assessment guidelines and expectations, can go a long way in promoting academic integrity and minimising the cases of academic misconduct.

3.3. Relevance

Professional relevance is crucial for effective pedagogy as it ensures that students acquire skills and knowledge that are applicable to real-world contexts (e.g. Teichler, 2015). As governments, businesses, and community organisations integrate generative AI tools into their workflows, ethical considerations, content verification, and responsible use these tools are becoming essential skills. Student guidelines on the use of generative AI, including ChatGPT, should align with industry standards to ensure they remain relevant and prepare students for their future careers. Moreover, we can safely assume that students will be more inclined to comply with guidelines that they can see as directly relevant to their future employment, rather than simply as our personal preferences for how they should or should not engage with generative AI.

3.4. Transparency

Transparency is a crucial aspect of academic integrity and essential for assessment validity (e.g. Dawson *et al*, 2024). Thus, having developed and provided clear and relevant guidelines, in the context of a trusting learning environment, we should encourage and expect students to be transparent regarding their use of generative AI tools in their assignments, including ChatGPT. Students can be asked to provide a general description of their engagement with these tools in the production of their work, an explanation of how outputs generated by ChatGPT were incorporated into their assignments (e.g. used to structure, summarise, or brainstorm), including a list of prompts used to generate those outputs. These requirements promote a culture of academic integrity, make it easier for teachers to assess student learning, and contribute to develop a critical and ethical disposition towards the use generative AI.

4. The Principles at Work - ChatGPT Appendix

In my teaching practice, these principles translate into an exploration of the pros and cons of ChatGPT (with the emphasis on caution), the provision of clear guidelines and the rationale for those guidelines, with references to university policies and industry standards (to demonstrate relevance), and the requirement to include a ChatGPT/GenAI Appendix when students use generative AI in the production of their essays (to encourage transparency). The whole approach is predicated on developing trust between students and teachers -notwithstanding that academic integrity requires that teachers 'trust but verify' if there are indications of academic misconduct.

4.1. Assessment Instructions

Having cautioned students about the bullshit artist nature of ChatGPT, outlined the relevant university policies and provided examples of industry standards, students are given clear and explicit instructions regarding the use of generative AI tools in their essays. I tell students that they can use generative AI (including ChatGPT) to assist with expression and idea generation,

but I remind them that the essay they submit must be their work. I also remind them that they are expected to fully understand every aspect of their essay, and that if there is a concern about the use of AI tools exceeding the assessment guidelines, they may be asked to discuss the assignment before the mark is finalised. I also advise them to keep a folder with drafts, notes, annotated readings, and a dot-point list of how their essays were produced, in case their authorship is questioned. Finally, I ask students to be transparent in their use of generative AI and include a ChatGPT/GenAI Appendix that explains how they have used these tools, including a list of prompts, if relevant. I communicate this instruction using the advice provided by Australian Customs and Biosecurity to people who arrive in Australia: “Don’t Be Sorry, Just Declare It.” Students are reminded of this requirement in the essay submission portal:

Don’t forget to include a ChatGPT/GenAI Appendix after the Reference List if you have used ChatGPT/GenAI in the production of the essay. The absence of this Appendix is equivalent to stating: “I did NOT use ChatGPT/GenAI”. If this statement turns out to be false, this would constitute a breach of Academic Integrity. Remember: “Don't Be Sorry, Just Declare It”.

4.2. Evidence of Impact

The evidence from the implementation of this initiative in two consecutive editions of the same course, with a combined enrolment of 214 students (113 in Semester 1, 2023 + 101 in Semester 2 2024), indicates that students respond well to the approach, and suggests that this approach can minimise the cases of academic misconduct. The approach was applied to the production of essays, which constituted 50% of the assessment. The assessment also included two in-person assignments: weekly seminar work (20%) and a multiple-choice test (30%). These in-person assignments help get a sense of the level of student understanding of the course content, which can inform academic judgement when there are concerns with student authorship of their essays.

The first time I used this approach, in 2023, most students chose not to use ChatGPT and those who used it did so in limited and acceptable ways. There were only three essays (out of 103 submissions) that required a follow up conversation. The second time, in 2024, once again, most students chose not to use ChatGPT, and those who used it did so in limited and acceptable ways. There was only one essay (out of 96 submissions) that required a follow up conversation.

Here is the breakdown of the 199 essays submitted based on their Turnitin AI score, which indicates the percentage of text within the submission that was likely generated using AI:

AI score of 0: 155 (78%). This includes cases of students who declared the use of ChatGPT, but who used it in perfectly valid ways, and thus it was not identified as an issue.

AI score of 1-20: 28 (14%). These were essays within the threshold for higher likelihood of false positives and thus scores that should not be pursued for academic misconduct; not to mention that no academic integrity issues were identified during the marking process.

AI score of 21-60: 12 (6%). These were cases that required closer examination, but ultimately none of them indicated a problematic use of generative AI and, importantly, the quality of these essays was reflective of the overall performance of those students in the course.

AI score over 60: 4 (2%). These were cases that required a follow up conversation, and the only ones that reflected a problematic use of generative AI. In two cases the issue was what I called “life”. Health, work and family issues led these students to use ChatGPT to cut corners, and this was obvious in the content and narrative style of their essays. These were not good essays. The essay marks were downgraded, and the students ended with grades that reflected their overall performance in the course: Fail and Pass. In the other two cases, the issue was what I called “voice”. The students run the essay content so many times, in different ways, through generative AI tools, particularly Grammarly (which not all students know is AI-powered), that by the end their “voice” had become robotic. These were students who did not think they were doing anything wrong and who ultimately did not benefit from the use of generative AI. The students ended with grades that reflected their overall performance in the course: Pass and Credit.

The evidence indicates that students respond well to this pedagogical initiative and reveals that this approach can minimise the potential cases of academic misconduct. In the first iteration, this approach reduced the cases of academic misconduct to a level that was barely noticeable. In the second iteration, which was more explicit and systematic, there were no cases of academic misconduct. This initiative has also been popular with colleagues, as evidenced by the numerous invitations to present this approach at learning and teaching seminars, communities of practice sessions, and professional development events, both within and beyond my own university.

5. Conclusion

The initiative outlined here to address the pedagogical challenges posed by generative AI (particularly ChatGPT) in higher education, especially when it comes to written assignments such as essays, does not require (although it is compatible with) redesigning those assignments. Instead, this initiative relies on building a trusting learning environment that encourages students to develop a critical and responsible approach to this revolutionary technology. The practice consists in: (1) explaining to the students the need to be cautious when using generative AI; (2) providing them with the rationale for the guidelines regarding the use of generative AI in their assignments, including how those guidelines reflect industry standards; and (3) requiring students to submit a ChatGPT/GenAI Appendix with their written assignments. The evidence suggests that an approach built on the principles of caution, trust, relevance and transparency can go a long way in addressing some of the most urgent challenges posed by generative AI (including ChatGPT), particularly concerns with academic integrity. This approach is neatly captured (and has been greatly enhanced) by the slogan: “Don't Be Sorry, Just Declare It”.

References

- Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 49(6), 893–905. <https://doi.org/10.1080/02602938.2024.2335321>
- Bobula, M. (2024). Generative Artificial Intelligence (AI) in higher education: a comprehensive review of challenges, opportunities, and implications. *Journal of Learning Development in Higher Education*, Issue 30, 1–27. <https://doi.org/10.47408/jldhe.vi30.1137>
- Bormann, I., Niedlich, S., & Würbel, I. (2021). Trust in Educational Settings: What It Is and Why It Matters. European Perspectives. *European Education*, 53(3–4), 121–136. <https://doi.org/10.1080/10564934.2022.2080564>
- Curzon-Hobson, A. (2002). A Pedagogy of Trust in Higher Learning. *Teaching in Higher Education*, 7(3), 265–276. <https://doi.org/10.1080/13562510220144770>
- Davies, S., Sormunen, K., & Kangas, K. (2024). Cultivating Epistemic Doubt: a Key Competence for Productive Participation in the Era of Artificial Intelligence. In Clarke-Midura, J., Kollar, I., Gu, X., & D'Angelo, C. (Eds.), *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning*, 415–416. International Society of the Learning Sciences. <https://doi.org/10.22318/cscl2024.114185>
- Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005–1016. <https://doi.org/10.1080/02602938.2024.2386662>
- Frankfurt, H. G. (2005). *On Bullshit*. Princeton, N.J.: Princeton University Press.
- Hicks, M. T., Humphries, J. & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, 26: 38, 1–10. <https://doi.org/10.1007/s10676-024-09775-5>
- Kaur, K. & Shilpa, R. J. (2024). The ChatGPT revolution: Reshaping education through artificial intelligence. *AIP Conference Proceedings*, 3121(1), 1–16. <https://doi.org/10.1063/5.0221430>
- Kurban, C. F. & Şahin, M. (2024). *The Impact of ChatGPT on Higher Education: Exploring the AI Revolution*. Leeds: Emerald Publishing Limited. <https://doi.org/10.1108/9781837976478>
- Luo (Jess), J. (2024). How does GenAI affect trust in teacher-student relationships? Insights from students' assessment experiences. *Teaching in Higher Education*, 1–16. <https://doi.org/10.1080/13562517.2024.2341005>
- Messeri, L. (2023). Teaching with ChatGPT: Critiquing Generative Artificial Intelligence from the Classroom. *Anthropology Now*, 15(1), 84–92. <https://doi.org/10.1080/19428200.2023.2230098>
- Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwing, N. & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 1–10. <https://doi.org/10.1016/j.caeai.2022.100075>
- Teichler, U. (2015). Changing Perspectives: The Professional Relevance of Higher Education on the Way Towards the Highly-Educated Society. *European Journal of Education*, 50(4), 461–477. <https://doi.org/10.1111/ejed.12146>