INNOVATION IN LANGUAGE LEARNING International Conference

Addressing the Dark Sides of AI for Teaching Critical Thinking Skills to Undergraduate Students

Dietmar Tatzl

FH JOANNEUM University of Applied Sciences, Austria

Abstract

This contribution casts a critical perspective on the overwhelming spread of generative artificial intelligence (AI) tools and applications throughout society. This AI hype has gripped educational institutions, business leaders, and political decision makers alike, so that voicing scepticism, concern, or disbelief almost amounts to heresy, and AI critics are faced with accusations of backwardness, technophobia, or conservatism. Yet it is more important than ever to warn of the dark sides of AI already known and potential future threats more difficult to discern. Higher education in particular has a responsibility to question the blind and naïve embracement of AI as the saviour of modern work life. In the end, university graduates need to be independent professionals with critical thinking skills to form, develop, and protect democratic societies. It is the aim of this contribution to present a way of honing undergraduate students' critical thinking skills by drawing their attention to negative aspects of AI and discussing its impacts on society, in the sense that students are trained in estimating the consequences of technological innovation on democracy, private life, and society.

Keywords: *AI, technology, society, critical thinking, ethics, undergraduate*

1. Introduction

Generative artificial intelligence (AI) is a technological breakthrough in computer science which is praised by its proponents as an all-pervasive digital revolution [1]. Since the release of OpenAI's ChatGPT in November 2022, the narrative has been forming that AI will lead to a boost in productivity, efficiency, and quality of life by freeing humans from time-consuming analytical tasks and thus unlocking their potential and working time for increasing the output of their activities. However, apart from resting on the false assumption that AI delivers faultless results, this modern fairy tale has also raised suspicions and concerns from critical minds from all spheres of life.

Researchers and academics from a variety of disciplines have been warning of risks and threats caused by AI. Klockmann et al. [2], for instance, argue that the elite groups of developers and those who train AI are not affected by the outcomes of their work, which becomes particularly dangerous for future generations who will be increasingly dependent on decision-making AI agents. Warnings have also been voiced from this inner circle of programmers and developers working on AI. Spiegler et al. [3] have identified three major concerns of AI practitioners they interviewed: harmful recommendations given by AI companions, an increasingly distorted reality caused by deep fakes and misinformation, and widespread job loss. Several high-ranking AI experts have recently resigned from their top positions at tech companies over safety concerns about the speed and uncertainty involved in the technology's development [4]. At the same time, organisations are trying to burden employees with the impossible task of applying AI responsibly. This demand for responsible use of AI is becoming a utopian ideal, particularly in the face of systemic risks resulting from the "complexity and poorly understood system interactions between humans, machine, and ecology" [5, p. 8]. Moreover, staff cannot responsibly apply a tool over which they only have limited control. The more intensely humans work with AI, the more dependent and vulnerable they will become.

In legal terms, Chatterjee et al. [6, p. 201] associate AI with "various entangled challenges jeopardising Human Rights". For instance, there is the danger of discriminating against disadvantaged groups of people, which becomes particularly precarious in healthcare. Fiske et al. [7, p. e292] address the dilemma of collecting race and ethnicity information in medical contexts for reducing bias in data on which AI relies but at the same time risking the abuse of such personal data to further marginalise already disadvantaged groups in healthcare systems. Walker et al. [8] have pointed to such algorithmic bias in AI as a crucial problem stemming from inequalities in power, which is detrimental to the automation of background processes in healthcare systems for the benefit of all patients. Legally, there are attempts to shift the focus from risk assessment to harm profiles of AI

systems, with the goal of enabling redress and compensation for harm caused by AI applications [9]. Success of such endeavours would be a welcome step in the right direction and could potentially protect people from AI abuse or at least equip victims with the means to take legal action.

Freitas et al. [10] point to ethical issues of AI use related to publication integrity and accuracy. Similarly, Hu [11] addresses the increasing number of fake papers and paper mills in the publication process, endangering the integrity and reputation of journals, editors, reviewers, and authors; in other words, undermining the whole scientific process. Retractions of papers based on the misuse of AI are likely to multiply in response to these developments. A recent example of such a retraction notice has been published by the Editors-in-Chief [12] of the *Journal of English for Academic Purposes*.

In educational policy, Wang [13] warns of confusing AI-generated recommendations with decisions and calls for transparency and the need to employ AI merely to support human decision-making. In fact, this is a general concern breaking disciplinary boundaries and affecting humans of all age ranges, although children, teenagers, and the elderly are particularly vulnerable groups. Cybercrime, terrorism, and disinformation campaigns by authoritarian regimes exacerbate this danger of influencing humans through AI to incite them to make harmful decisions.

Another major threat of AI stems from this technology's reliance on massive computing power. Gross [14] underscores the tremendous energy consumption and carbon emissions caused by AI use, which fuels the climate crisis. Although this existential threat is well known, AI's exact impact on the climate still needs to be determined.

This list of harmful effects of AI from various areas of life is far from comprehensive, but it should suffice as an overview of potential topics for in-class discussions, as this paper introduces a pedagogical stream to teach critical thinking skills to undergraduate students.

2. Teaching Critical Thinking Skills in the Local Framework

In response to the advent of AI, it was deemed necessary to address the negative aspects of this technology and its impacts on society in class. Higher education has a responsibility to equip its graduates with critical thinking skills in their professional fields and beyond to form, develop, and protect democratic societies. Students in this teaching stream are enrolled in an undergraduate engineering program in Austria. As part of their studies, they take English for specific purposes (ESP) courses to improve their technical language skills. This contribution presents a teaching stream that is integrated in these ESP courses and aims at questioning the narrative of AI as the saviour of modern work life. It is its goal to sharpen undergraduate students' critical thinking skills by rejecting the naïve embracement of AI as just a tool for improving work processes or planning activities. Through this stream, students are trained in assessing the consequences of technological change and innovation on their professional fields and on broader society.

2.1 Discussion Activities

The core of this teaching stream consists of discussion activities based on short quotations on AI and automation. Each of these activities is announced by the heading "AI criticism from 2024 Nobel prize winners in economics" and builds on extracts from Acemoglu and Johnson [15]. In this undergraduate program, ESP courses are interspersed with between two and six AI discussions over a whole semester to regularly practise critical thinking skills. About ten minutes are reserved for each of these, but if groups are in a talkative mood or gripped by the focus of the ongoing debate, there is flexibility to expand this time limit. The group size also varies from three to four participants. The discussions are introduced by the following instructions: *Reflect on the quotation below in small groups. Discuss current and potential future dark sides of AI. Exchange critical views from the media and personal experience with AI use*. While learners are commenting on the quotation and contributing their viewpoints, the teacher circulates and may steer the conversations in a certain direction, depending on students' inputs.

What follows is a selection of quotations employed in this teaching stream. Each of these extracts criticises different aspects of AI and their effects on society. This list here serves the main purpose of illustrating the function of the quotations as discussion triggers in the activities. The phrases at the beginning summarise the extracts' main points of criticism, although the debates should not be limited to these but rather be inspired by them to include further related risks and threats of AI. Students are expected to both respond to the quotations and exchange their own experiences with AI. In this way, each extract delivers a starting point for critical thinking and invites reflection on personal

encounters with this technology. Apart from honing their critical thinking skills, students practise communicative interaction, argumentation, oral fluency, and technology assessment. A further objective of these extracts is to raise learners' interest in the whole book [16] and other publications critical of AI for independent reading, so that they may find merit in continuing their engagement with the topic and in broadening their understanding of the interrelationships between technology and society.

AI for surveillance and control of citizens

> Under the leadership of Xi Jinping, the government increased its demand for surveillance and related AI technologies first in Xinjiang and then throughout China. In 2017 it issued the "New Generation AI Development Plan," with a goal of global leadership in AI and a clear focus on the use of AI for surveillance. Since 2014, China's spending on surveillance software and cameras and its share of global investment in AI have increased rapidly every year, now making up about 20 percent of worldwide AI spending. Researchers located in China now account for more AI-related patents than any other country. [17, pp. 345–346]

AI is amplifying antidemocratic forces

> The AI illusion thus favors an antidemocratic impulse, even as many of its executives view themselves to be on the center-left and supporters of democratic institutions and even the Democratic Party. Their support is often rooted in cultural issues and conveniently bypasses the vital building block of democracy: people's active participation in politics. Such participation is especially discouraged when it comes to AI because most entrepreneurs and venture capitalists believe that people do not understand the technology and unnecessarily worry about its intrusive effects. [18, pp. 372–373]

AI for monitoring workers and polarising society

> The tragedy is that AI is further undermining democracy when we need it most. Unless the direction of digital technologies is altered fundamentally, they will continue to fuel inequality and marginalize large segments of the labor force, both in the West and increasingly around the world. AI technologies are also being used to more intensively monitor workers and, through this channel, create even more downward pressure on wages. [19, p. 380]

AI for enriching a few but failing to share gains with broader society

> The marriage of digital technologies and big business had created a growing number of billionaires by the mid-2000s. Such fortunes multiplied once AI tools started spreading in the 2010s. But this was not because AI turned out to be anything as productive or amazing as its boosters have maintained. On the contrary, AI-based automation often fails to increase productivity by that much. Worse, it is no way to build shared prosperity. It nevertheless enthralls and enriches tycoons and top managers as it disempowers workers and opens up new ways of monetizing information about people […]. [20, p. 337]

AI moving in the wrong direction

> Our current problems are rooted in the enormous economic, political, and social power of corporations, especially in the tech industry. The concentrated power of business undercuts shared prosperity because it limits the sharing of gains from technological change. But its most pernicious impact is via the direction of technology, which is moving excessively toward automation, surveillance, data collection, and advertising. [21, pp. 392–393]

2.2 TED Talks

Two TED talks on AI and algorithms accompany the work with text extracts. The first one is an early promotion of AI as a paradigm shift in computer science and economics by the speaker, who explains

how this technology may affect industry in the future [22]. It also invites a comparison of the predictions made in 2016 with the current state of AI development and implementation. Students are asked to take notes during the talk and discuss these questions after having watched the video: *Would you agree with the speaker that AI equals an industrial revolution? Which areas of human activity will AI outperform and maybe replace?* The group size here varies between two and four.

The second TED talk [23] addresses algorithms that filter online information. Students again take notes on the content of the talk, and, afterwards, answer these questions in small groups: *Why is filtered information problematic? What can we do to escape from the filter bubble? Where can we find reliable and objective information?* This topic is well suited to incite discussion and blends with the core stream of AI criticism, as it centres on algorithmic bias and the patronising principle underlying personalised search results.

3. A Note of Caution for Students

The critical thinking stream in the author's undergraduate program is complemented by a note of caution in students' course materials. This note warns learners of certain risks involved in AI tools and provides some guidance for their use. The note points to the nature of AI and intends to prevent course participants from endangering their own and others' digital safety and security. Figure 1 shows the text as it appears in the course materials.

Be aware that generative AI tools serve the main purpose of collecting massive data from users to further feed, develop, and train algorithms. Use generative AI with care and do not upload or allow access to:

- Published content protected by copyright law
- Confidential material protected by non-disclosure agreements with companies
- Personal data, documents, and photos
- Comments spreading discrimination, hate, and violence
- Fake news and disinformation
- Material threatening the dignity, liberty, and life of others
- Questions that reveal or compromise, by inference, any of the points listed
- AI agents such as OpenClaw requesting control of your data and user rights

Fig. 1. Note of caution warning students of dangers inherent in AI tools

4. Discussion and Recommendations

The teaching stream on critical thinking presented here resulted in many lively discussions among students. Critical inputs from participants on their experiences with AI enriched the quotations used as triggers of debates and expanded the range of aspects which the groups treated. As AI is ubiquitous in the media all over the world, taking it as the thematic core for teaching critical thinking skills facilitated students' active participation in the discussions. However, the teaching stream can be diversified by adding quotations from related fields for the group activities, which may also reduce the risk of boring learners with the same topic in several lessons. Such a reaction of tiring has not yet been observed in class, as AI developments are rapid and new aspects of the technology surface almost weekly, but it may be reasonable to prepare a broader spectrum of discussion topics to remain flexible as a teacher.

Educators in similar undergraduate programs may find it necessary to teach critical thinking skills to their students, and it is hoped that the activities, questions, and quotations presented here will support their endeavours. The teaching stream introduced may be replicated in undergraduate settings around the world or serve as inspiration for further development. It may be used in parts or as a whole, but it should be tailored to local educational situations to maximise its pedagogical benefits.

5. Conclusions

Adapting to the new realities caused by the widespread integration of AI in technology may be an enormous challenge for humankind. One needs to keep in mind that AI is already a faulty generative tool and even less reliable for factual information gathering [24]. The task of using AI responsibly is already beyond human capabilities alone. Hamed et al. [25] call for computational fact-checking, transparency of algorithms, and the incorporation of comprehensive resources for result verification as ways of ensuring ethical AI operations and outcomes. Detecting and correcting the hallucinations of a chatbot, however, are merely fragments of the problems which may be created by AI in the future and

which humans will need to solve. The increasing installation of AI in electronic devices, household appliances, and cars, for instance, will further rob humans of choices and consent when faced with the difficulties of opting out of AI in certain products and services. The least we can do as language teachers is sharpening the critical thinking skills of students, so that they are better prepared to question the promised benefits of digital inventions, assess the impact of technology on society, and protect themselves and others from harmful and immature innovations.

REFERENCES

[1] Kelly, K. (2016). How AI can bring on a second Industrial Revolution [digital video file]. *Ted.com*. https://www.ted.com/talks/kevin_kelly_how_ai_can_bring_on_a_second_industrial_revolution

[2] Klockmann, V., von Schenk, A., & Villeval, M. C. (2022). Artificial intelligence, ethics, and intergenerational responsibility. *Journal of Economic Behavior and Organization, 203*, 284–317. https://doi.org/10.1016/j.jebo.2022.09.010

[3] Spiegler, S., Hoda, R., & Pant, A. (2026). Images of AI: How AI practitioners view the impact of artificial intelligence on society, now and in the future. *Technology in Society, 84*, 1–17, 103109. https://doi.org/10.1016/j.techsoc.2025.103109

[4] Morrow, A. (2026, February 11). *AI researchers are sounding the alarm on their way out the door*. CNN Business. Retrieved February 11, 2026 from https://edition.cnn.com/business

[5] Galaz, V., Centeno, M. A., Callahan, P. W., Causevic, A., Patterson, T., Brass, I., Baum, S., Farber, D., Fischer, J., Garcia, D., McPhearson, T., Jimenez, D., King, B., Larcey, P., & Levy, K. (2021). Artificial intelligence, systemic risks, and sustainability. *Technology in Society, 67*, 1–10, 101741. https://doi.org/10.1016/j.jdrv.2026.03.003

[6] Chatterjee, S., Sreenivasulu, N. S., & Hussain, Z. (2022). Evolution of artificial intelligence and its impact on human rights: From sociolegal perspective. *International Journal of Law and Management, 64*(2), 184–205. https://doi.org/10.1108/IJLMA-06-2021-0156

[7] Fiske, A., Blacker, S., Geneviève, L. D., Willem, T., Fritzsche, M.-C., Buyx, A., Celi, L. A., & McLennan, S. (2025). Weighing the benefits and risks of collecting race and ethnicity data in clinical settings for medical artificial intelligence. *The Lancet Digital Health, 7*(4), e286–294. www.thelancet.com/digital-health

[8] Walker, R., Dillard-Wright, J., & Iradukunda, F. (2023). Algorithmic bias in artificial intelligence is a problem—And the root issue is power. *Nursing Outlook, 71*, 1–4, 102023. https://doi.org/10.1016/j.outlook.2023.102023

[9] Li, S., & Heine, K. (2026). Developing a harm-based approach to understand digital vulnerability in the era of AI: A perspective of the European Union. *Computer Law & Security Review: The International Journal of Technology Law and Practice, 60*, 1–13, 106266. https://doi.org/10.1016/j.clsr.2026.106266

[10] Freitas, G., Ünal, I. H., Goldust, M., & Grant-Kels, J. M. (2026). Challenges and ethical issues of maintaining publication integrity and accuracy with the use of artificial intelligence. *JAAD Reviews, 8*, 63–68. https://doi.org/10.1016/j.jdrv.2026.03.003

[11] Hu, G. (2026). The crushing weight of fake papers. *Journal of English for Academic Purposes, 81*, 1–2, 101688. https://doi.org/10.1016/j.jeap.2026.101688

[12] Editors-in-Chief. (2026). Retraction notice to “Guided and unguided GenAI tasks for learning academic spoken collocations” [Journal of English for Academic Purposes 80 (2026) 101638] by V. Morgana & F. Poli. *Journal of English for Academic Purposes, 82*, 101735. https://doi.org/10.1016/j.jeap.2026.101735

[13] Wang, Y. (2021). When artificial intelligence meets educational leaders’ data-informed decision-making: A cautionary tale. *Studies in Educational Evaluation, 69*, 1–9, 100872. https://doi.org/10.1016/j.stueduc.2020.100872

[14] Gross, M. (2024). The artificial intelligence tipping point. *Current Biology, 34*(15), R709–R711. URL: https://www.sciencedirect.com/journal/current-biology/vol/34/issue/15

[15] Acemoglu, D., & Johnson, S. (2024). *Power and progress: Our thousand-year struggle over technology and prosperity* (Pbk. ed.), London, UK, Basic Books.

[16] Acemoglu & Johnson (2024).

[17] Acemoglu & Johnson (2024).

[18] Acemoglu & Johnson (2024).

[19] Acemoglu & Johnson (2024).

[20] Acemoglu & Johnson (2024).

[21] Acemoglu & Johnson (2024).
[22] Kelly (2016).
[23] Pariser, E. (2011). Beware online “filter bubbles” [digital video file]. *Ted.com*. https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles/up-next#t-525120
[24] Morrow, A., & Eadicicco, L. (2025, August 14). *The latest ChatGPT is supposed to be ‘PhD level’ smart. It can’t even label a map*. CNN Business. Retrieved August 14, 2025 from https://edition.cnn.com/business
[25] Hamed, A. A., Zachara-Szymanska, M., & Wu, X. (2024). Safeguarding authenticity for mitigating the harms of generative AI: Issues, research agenda, and policies for detection, fact-checking, and ethical AI. *iScience, 27*, 1–6, 108782. https://doi.org/10.1016/j.isci.2024.108782