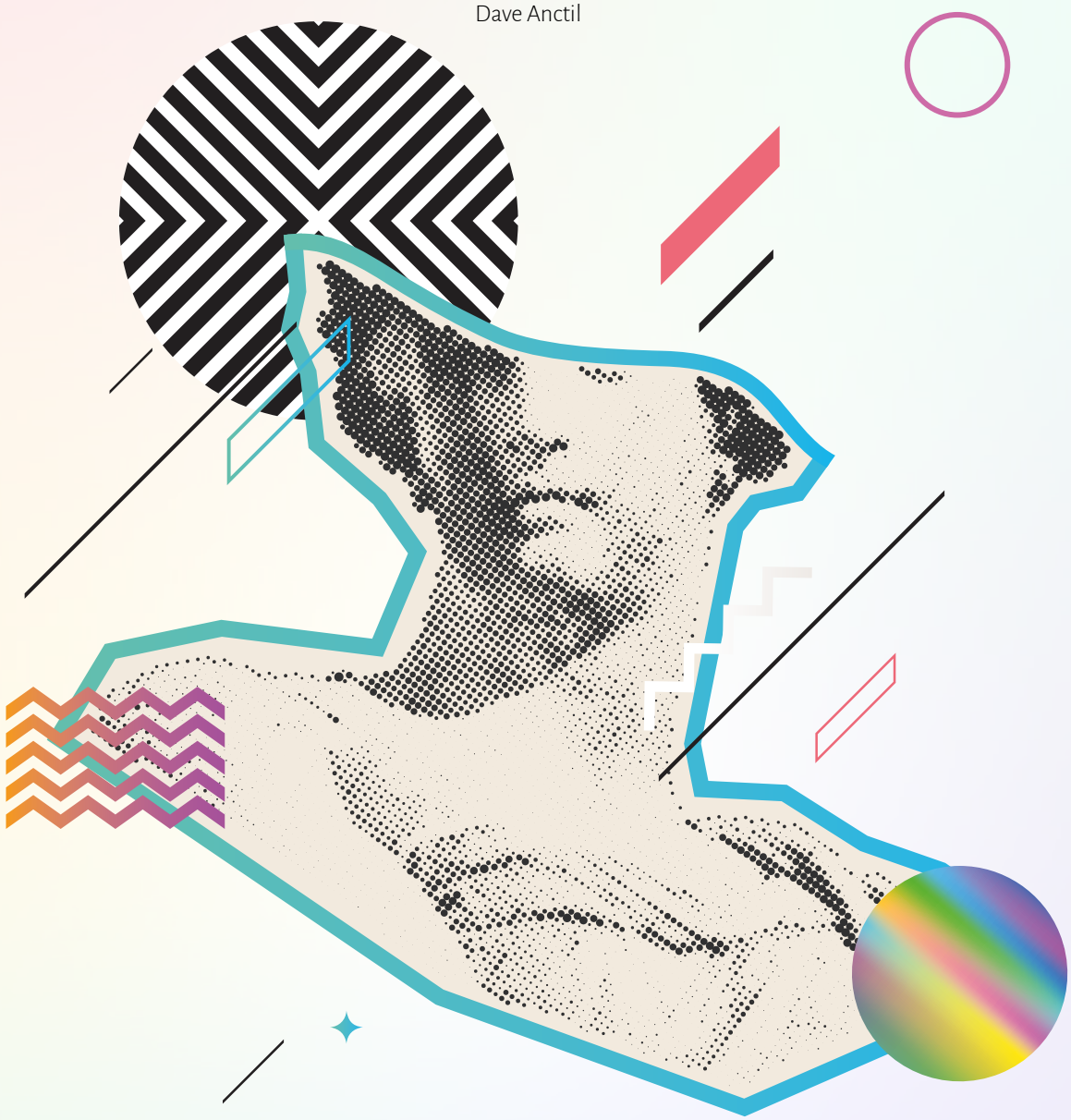
Higher Education in the Era of Generative AI

Dave Anctil



The year 2023 will have brought the education community to realize the state of artificial intelligence (AI) advancements with the ChatGPT app, which has become an instant global sensation. The higher education network now understands that AI tools can automatically write assignments and exams in all subjects and disciplines. Similar, though less powerful, tools have been around for several years. They were being used by a growing number of students, often without the knowledge of administrators and faculty, undermining intellectual integrity and academic justice. Now that the subject is in the media, it will no longer be possible to practise wilful ignorance about these practices that have been documented in academic integrity research.

The wide accessibility and increasing performance of new algorithmic systems are causing a loss of bearings and legitimate concerns. There is no doubt that AI will transform our profession in ways that are difficult to imagine. Information and communication technologies (ICT) have already imposed continuous adaptations among education professionals, especially since the turn of the 1990s-2000s. However, ICT has not fundamentally changed our professional role in the transmission and assessment of competencies and knowledge, nor has it changed the role and responsibilities of learners. Yet, it is partly these traditional roles and conceptions of teaching that generative AI is disrupting.

This article first provides a brief, popularized overview of generative AI technology. In light of research on academic integrity and on the digital practices of the student community, it then uncovers the key issues that should capture our attention and guide our action. Finally, I would like to offer my perspective in the broader debate on the integration of AI technologies in higher education.

Generative AI

Automatic content generation technology (or "generative AI") has been around for several years. If this technology has become a major news item in 2023, it is because of the ChatGPT application, made available for free to the general public on November 30, 2022 by the company OpenAI.[1] What followed was the beginning of an accelerated race between technology companies that are now vying to perfect these systems and make them easier to access, especially through applications like ChatGPT, but also Google and Microsoft browsers and search engines, and soon the Microsoft 365 suite. The number of new generative AI applications available already numbers in the thousands.[2]

What is ChatGPT? The user interface of a chatbot. It allows an individual with no particular technical skills to generate text through "prompts," i.e. questions and suggestions written in "natural language," such as French and English, or a mixture of languages, such as computer code or various mathematical and scientific languages. The latest applications even allow for the input of images, graphics or files in various formats.

[1] Readers interested in learning more about the language model or testing the ChatGPT application can visit [openai.com/blog/chatgpt].

[2] The thousands of AI applications available to the public, either free or by subscription, are now listed and ranked by users on Futurepedia - The Largest AI Tools Directory [futurepedia].

New generative AI systems do not process information like a conventional search engine, such as a classification algorithm based on statistical association of keywords to search a database or the Internet. Instead, the inputs are processed by a mathematical system that is much more powerful than search engines. This system was developed to understand the meaning and context of a query in order to statistically predict the best possible answers, building on its own internal knowledge. For each query, the system generates original output, which can be refined or enhanced automatically with other prompts: "reduce the text by half," "add two examples," "apply a function," etc. ChatGPT and Playground (OpenAI), Bard and LaMDA (Google) or Bing AI (Microsoft and OpenAI) are only the most successful and popular examples of these systems.

Before explaining further, let's first note that natural language processing (NLP) systems have existed for a long time. Different methods allow computers to process languages. We have been using them for several years for speech recognition, text classification and analysis, automatic translation and correction, sentiment analysis and of course content generation. For generative AI, it is important to recognize that it is not the conversational interface that provides computers with the ability to understand language and generate content, but rather what AI researchers call a "generative model" of AI. This is a very powerful piece of software, developed by machine learning on specialized supercomputers at a cost of millions of dollars, and then housed on very large networks of servers that provide access to the model via applications. For example, the GPT-3.5 model (for generative pre-trained transformer version 3.5) is a particular model, but it powers several different interfaces and applications available online.

The recent explosion in the performance of generative models was made possible by major innovations in the field of AI called "deep learning," which date from 2017. They have allowed the development of "large language models" (LLMs). And these are the models that demonstrate unprecedented capabilities, notably by breaking records in standardized AI performance tests (AI benchmarks), but also by passing academic and professional exams (more on this later).

It is difficult to explain how large language models work without using technical terminology. These models differ from other AI systems because they are actually capable of transforming training data into organized knowledge, not just assisting a user with a search algorithm. This difference is crucial. It sets deep learning apart from other algorithms and approaches to AI that are more familiar to the general public.

Indeed, language models are built on a network composed of billions of artificial neurons, whose organization mathematically imitates the biological functioning of neurons and synapses in the brain. During the training phase, which feeds this neural network with billions of data of various formats (training datasets), a programmed architecture allows the network to organize its terabytes of information into billions of objects and statistical relations between these objects, in the form of mathematical vectors. The model is then tested and refined using a learning method based on human feedback called "RLHF" (for reinforcement learning from human feedback). This makes it possible to correct structural errors in the model according to the goals pursued by the engineer-developers.

A generative LLM application can thus perform all of the above tasks, previously performed by separate, specialized AI systems, at an ever-increasing level of performance. In this way, AI engineers produce ever more powerful models, and do so by increasing the size of the neural network, the quality of the training data, as well as the refinement of the RLHF methods. LLM applications can now not only generate text, but also code, solve mathematical and scientific problems, and elaborate assignments in most technical and non-technical fields. They are even beginning to generate music and video (OpenAI, 2023; Google Research, 2023).

Image generators, such as DALL-E (OpenAI) and Midjourney (independent laboratory), have been developed as separate models over the past few years. However, in 2023, text, image and audiovisual databases are gradually being merged into new multimodal models, i.e. a single model will now be able to generate all types of content. This is notably the case with the GPT-4 model (OpenAI and Microsoft), which is also capable of creating applications and websites.

The explosion in the performance of generative AI applications means that it will now be possible, as early as 2023, to entrust them with most of the cognitive tasks taught and assessed in higher education, from textual or visual, audio or audiovisual artistic

creation to philosophical, mathematical and scientific reasoning, as well as programming or strategic planning (OpenAI, 2023).

I am fully aware that the previous statement about the artificial cognition of generative AI will seem surprising to non-specialists. However, it is corroborated by dozens of research labs that specialize in empirically evaluating the progression of cognitive capabilities of AI systems (Wei *et al.*, 2022). As an example: although the GPT-4 model was released on March 14, 2023, a preliminary version was tested by researchers at Microsoft Research. Their evaluation report is based on dozens of standardized tests to measure and compare human cognition with that of the models. It shows that in all the tasks measured, the performance of GPT-4 is close to or better than human performance, and far exceeds the previous GPT-3.5 model (Bubek, 2023). It is important to ponder the significance of the tentative conclusion of this first large-scale research on the cognition of large language models like GPT-4:

> The central claim of our work is that GPT-4 attains a form of general intelligence, indeed showing sparks of artificial general intelligence. This is demonstrated by its core mental capabilities (such as reasoning, creativity, and deduction), its range of topics on which it has gained expertise (such as literature, medicine, and coding), and the variety of tasks it is able to perform (e.g., playing games, using tools, explaining itself, ...). (Bubek, 2023, p. 92)

There is no doubt that AI will transform our profession in ways that are difficult to imagine.

Other independent laboratories are currently evaluating the surprising capabilities of GPT-4, and we will soon learn more about its extent.

Despite the research, some people find it hard to imagine that AI systems can do more than memorize and repeat the output of human cognition. But this is not the case. Large models like GPT-4 do not have human cultural productions in memory in a database. Rather, as explained above, they are large-scale models of knowledge organized mathematically through deep learning techniques. Although trained on human intellectual and cultural productions, they do generate their own cognitive processes in their content productions. The human formatting of written queries (prompting) can however lead the model to imitate human styles, for example by composing an illustration "in the style of Picasso" or a poem or music "in the style of Baudelaire" or "in the style of Mozart." However, the model does not contain these works in a database; rather, it has assimilated information structures (tokens) of the works and texts that describe or discuss these works. In this sense, it does not "copy" these works, but generates an original production, based on its own understanding of these works, as human creators would do when inspired by the great masters. This also explains why experts in literature, visual arts or music composition will obtain much better results than non-experts: by specifying the details of the prompts based on their own knowledge of arts and sciences, the users of the language model increase (statistically) the quality of the model's prediction and, consequently, the conformity of the result obtained to the desired result.

There is currently no consensus on the level of intelligence of language models since there is no scientific consensus on the comparative intelligence of species in the sciences that study these issues. For example, non-human animals show very different forms of

intelligence in their structures and functioning than humans. Yet, these animals are capable of performing certain cognitive tasks at a level comparable or superior to that of humans. The comparison with AI systems is both simpler, since they can understand language—and thus pass human tests—and also more complex, since their cognitive processes are not biopsychological, and the interpretation of the functioning of the largest artificial neural networks is exceedingly difficult.

What the aforementioned studies measure, therefore, is the performance of models like GPT-4 in tasks normally performed by humans, and not the underlying cognitive processes involved in solving these tasks. Artificial (mathematical, computational) cognition is profoundly different from natural (biopsychological) cognition, and cognitive sciences applied to both kinds of cognition are just beginning to study the properties and functioning of large models developed in deep learning.

Returning to our central topic, higher education, we must recognize that the capabilities of the latest models of generative AI can now reach human-level intellectual production in a host of tasks, and that the progression of their cognitive capabilities has been growing exponentially over the past decade.

Issues for higher education[3]

The intellectual integrity policies of colleges and universities are based on the principle that students should do all the intellectual work in their evaluations. However, academic integrity research has shown that this principle has already been overtaken by the digital reality. As evidence of this, a 2009 study concluded that automatic academic fraud detection software tools would always lag behind the innovation and digital practices that allow them to be circumvented (Bretag & Mahmud, 2009). This prediction was abundantly confirmed.

Here are some concrete examples, well known to researchers. Digital tools for automatically paraphrasing plagiarized excerpts, such as Rephraser, have been used for at least five years to escape detector tracking. Other AI tools dramatically increase writing capabilities, such as Grammarly and InferKit. They allow mediocre writers to achieve high scores through assisted editing, proofreading, rewording, and completion of their work, without plagiarizing any sources.

When it comes to source plagiarism, it has been very easy to use machine translation to bypass the screening of teachers and detection software in recent years. The neural version of Google Translate (2016-) and DeepL Translate (2017-)—applications developed with deep learning—allow students seeking to plagiarize to copy-translate-paste chunks of text to make them undetectable with a bit of artisanal camouflage or by using Rephraser or Grammarly. The cheating student can even check the traceability of their fraudulent text with Turnitin and Google Search, often directly from these applications…

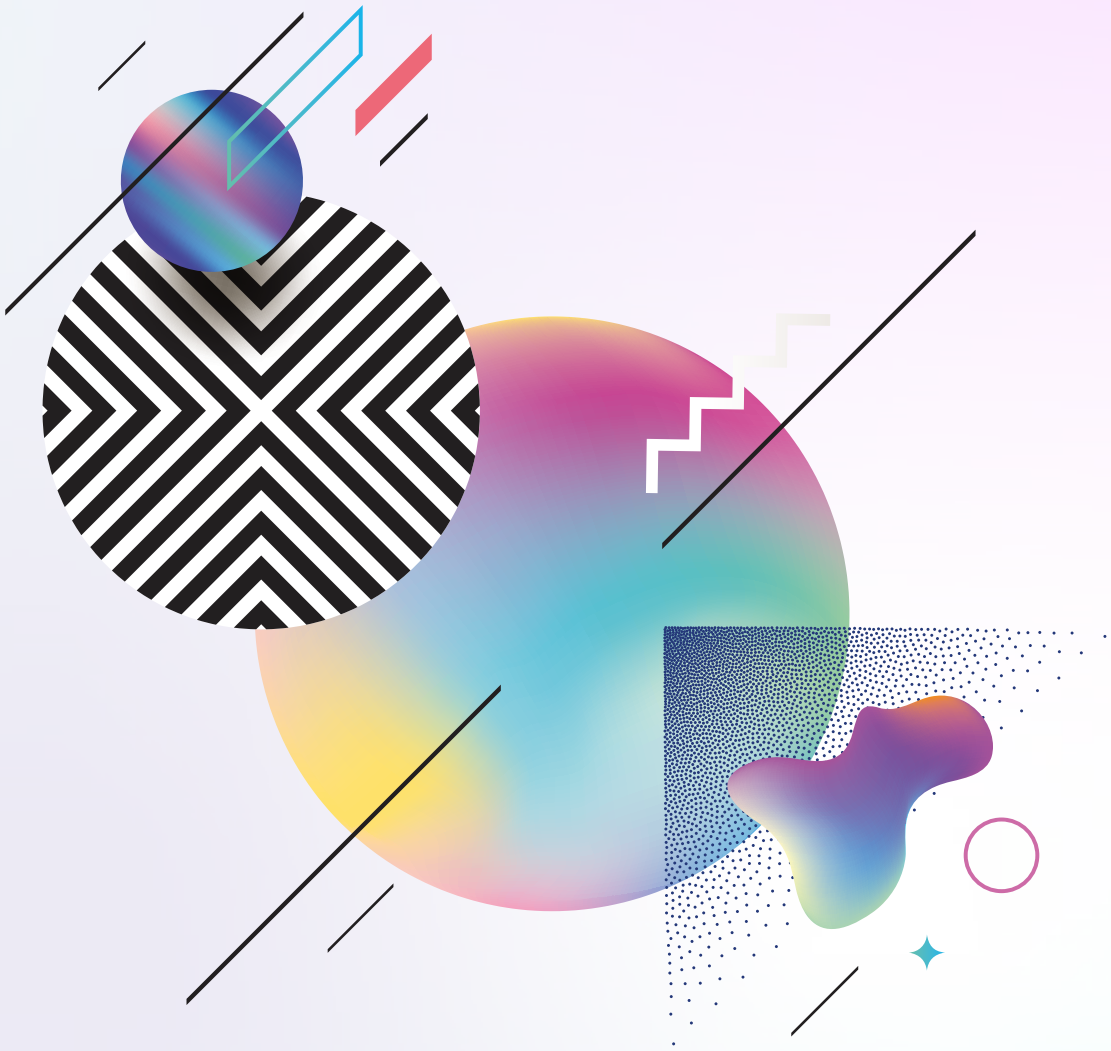
Recent research on integrity in the digital age should make us recognize the following reality: a significant number of students are already using a wide variety of AI tools to generate their assignments or to improve their academic performance with impunity, and have been for several years (Ahsan, Akbar & Kam, 2022; Surahman & Wang, 2022; Eaton & Hughes, 2022).

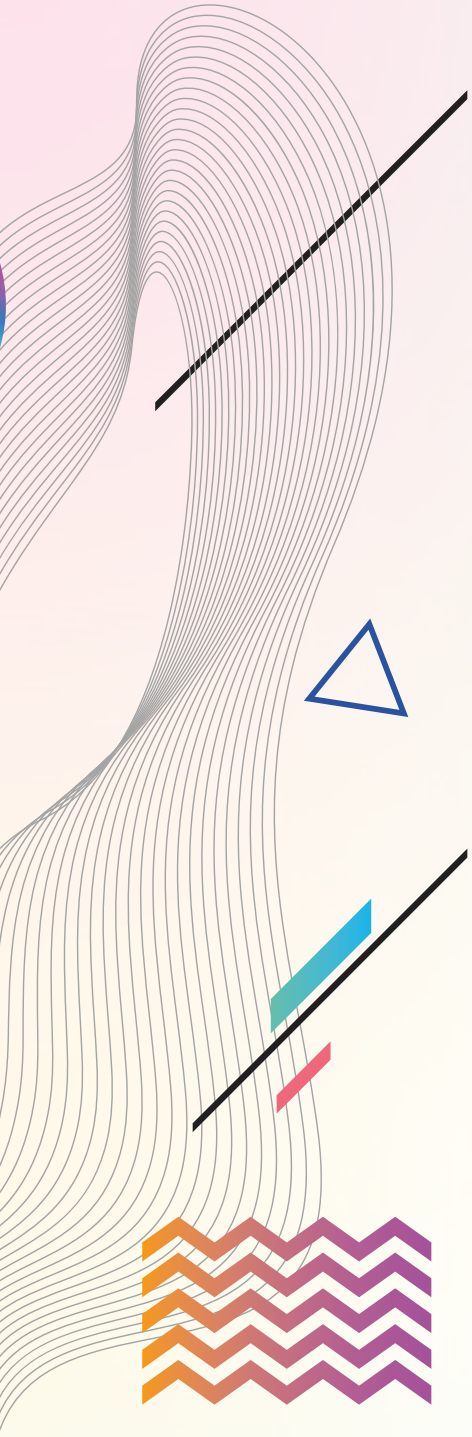
We should also recognize that it is not at all clear from the statements in academic policies that all of these practices are fraudulent or dishonest, since not all automatically generated content is plagiarized and the tools in question are rarely mentioned in the list of prohibited practices. To put it bluntly: higher education has not sufficiently recognized the challenge posed by the extremely rapid progression of digital technologies.

Therefore, in the face of ChatGPT, which has brought us abruptly into the era of generative AI, it is high time that CEGEPs and universities recognize that our intellectual integrity policies have long been out of step with technological and digital realities and student practices. In my opinion, adding new generative AI tools to the list of sanctioned prohibitions will not be enough to prevent dishonesty and develop intellectual integrity in the student community.

[3] I teach only in pre-university programs at the college level, as well as at the university level. Therefore, I am not qualified to reflect on issues at the elementary, secondary and technical college levels.

Moreover, looking for a "technical solution" to this problem is bound to fail. Contrary to popular belief, it is virtually impossible to detect the fraudulent use of new applications in academic work. The reason is simple: AI-generated content is original content comparable to that of an expert writer who draws on their sources. Despite announced plans for detector software, based on statistical methods, detection will never be able to offer a prediction beyond a reasonable doubt; it will always remain questionable. And fraud detection will, as in the past, be easily circumvented by following simple recipes available in the corners of the Internet—which is also true for the pseudo solution of digital watermarks.





It therefore seems urgent to us that CEGEP and university administrations and departments adopt two general and complementary approaches, summarized here before explaining them:

- **The avoidance approach**: protect the principles of assessment integrity and academic equality and equity by moving the most important evaluations measuring knowledge and competency attainment back into the controlled environment of the classroom now.

- **The integration approach**: quickly equip faculty with support to receive adequate training to 1) identify safe, ethical AI tools appropriate to their disciplines, and 2) explore the best ways to integrate these tools into teaching, pedagogical practices, and evaluations.

In the short term, the avoidance approach protects the integrity, equality and equity of academic evaluation. But it also has a cost: it could diminish the diversity of learning experiences offered to the new generation of students who, it should be remembered, have become accustomed to "project-based pedagogy" based on participatory evaluations that are more creative and complex than a simple exam (written or oral). Of course, modern higher education already has a wide variety of active evaluations, which are not limited to exams. For example, various collaborative evaluations and labs can be encouraged that do not require the use of digital technology, but that engage the individual and collective cognition of the learning community without the use of AI.

The second approach counterbalances the disadvantages of the first, by promoting the integration of generative AI tools into our educational philosophy and practices. This integration is naturally more popular with ICT pedagogical counsellors and technophile faculty members.[4] It consists of creating content and evaluation formats that encourage responsible and beneficial use of AI tools in terms of competencies and knowledge. It also has the merit of familiarizing educators and learners with how to use these tools with integrity, while also encouraging responsible use and developing critical thinking, in order to facilitate harmonious cooperation between human and artificial intelligences.

[4] I recommend that teachers empower themselves by consulting the many online resources, such as the Eductive site, that offer such pedagogical content. Yann Houry, Director of Pedagogical and Technological Innovation at the Lycée français international de Hong Kong, offers innovative and advanced experiments and technopedagogical video tutorials, notably from his LinkedIn page, which I warmly recommend.

Here is an example that could be called "the conversational feedback approach":

Learning to reason logically, including how to explain and argue well in writing, is one of the most important transversal competencies in higher education. But this broad competency is also difficult to acquire, since it requires a great deal of practice, and the teacher does not have enough time to individually supervise the grammatical, syntactic, and logical construction of the sentences, paragraphs, and texts of each individual in their groups. Applications of GPT-4 (OpenAI) or Claude+ (Anthropic) models can provide good feedback on different writing exercises. The idea is to provide exercises with pre-written prompts to encourage students to practise with the AI, in an autonomous and supervised way, rather than asking the system to do the work for them. The learner first writes explanatory and argumentative paragraphs based on the course content. Then they enter each text written in the dialog box.

Finally, they ask the AI prompts like:

> "Here is a reasoning paragraph [explanatory or argumentative on such and such a subject] that I have composed. I want to know the strengths and weaknesses of my text according to the [following criteria…]. I want you to explain my mistakes in detail, and give me suggestions for improving my reasoning…"

Conversational feedback is becoming a well-known form of tutoring assistance, which involves practising learning basics. It quickly leads the learner to see their strengths and weaknesses without fear of judgment or disruption. The teacher guides the learner toward a metacognitive use of AI. The learner then gets support with regard to the individual challenges of learning the tasks concerned, and practises better understanding their mistakes to improve. Teachers can then request to receive the result of this collaboration, in the form of a document where the learner reports their own progress, to assess effort and improvement before exams taken without assistance.

Space to develop other examples is lacking, but it should be understood that educational research on the proper uses of generative AI is still in its infancy. Scientific studies will be needed to know which practices are most likely to improve learning and cognition, while protecting the integrity of evaluations. Moreover, the integration of AI in education should not be furthered without asking philosophical and social questions about higher education in the age of cognitive AI.

The debate on AI integration in college

The perspective outlined so far is opposed to two dominant conceptions in the broad debate on the place of technology in higher education. It is opposed to a neo-liberal conception, oriented toward clientelism and blissful adaptation to the demands of the market (defined by the short-term needs of companies). But it is also opposed to a conception of education that could be called "romantic," and which consists in essentializing traditional educational practices by evoking the need for mobilizing teaching staff against the spectre of "technological dehumanization."

In order to better define my own perspective in the debate, I clarify these criticisms without, however, deepening them by taking into account all the nuances.

The neoliberal conception is not recommendable, not only because it betrays the fundamental principle of higher education as a public good,[5] but also because it is contradictory to the reality of technosocial evolution. Training students for the "reality" of the labour market seems reductionist and counterproductive, since the speed of transformation of the latter is constantly accelerating. It is therefore much wiser, as even the best economists specializing in this field admit, to offer a robust and versatile intellectual training that fosters the development of better cognitive, adaptive, and empowering capacities of students in the face of the socioprofessional challenges and transformations accelerated by AI (Brynjolfsson, 2023).

Conversely, the militant-romantic conception implies considering this technological progress as a "dehumanizing force" that should be "fought," defining education as a counterforce opposed to the destruction of culture, of what makes the human, the common, nature, life in general (Martin & Mussi, 2023). Although we subscribe to the Quebec humanist conception of education—to educate citizens and autonomous individuals capable of fully exercising their freedoms of thought and expression—the romantic discourse evoking "technical dehumanization" to reject digital technology and AI seems to serve militant political rather than educational interests.

Algorithmic technologies are redefining the human condition in the same way as other great inventions in history, such as writing and printing, the steam engine and electricity. To essentialize education by erecting its traditions as bulwarks in the face of technosocial change is to impose a single doctrine of the human being, for the benefit of a militant position in contradiction with the primary vocation of higher education: the free search for truth and the open transmission of knowledge. While critique of technologies must of course have its place in this search and transmission, we do not have the right to impose a particular ideology, with its biases and its intrinsic limits, under the pretext of "protecting" the student community against the technological civilization to which it belongs de facto and de jure.

As educators and specialists in our disciplines, our primary responsibility is to provide the best education possible, making choices based on our expertise, experience, and reliable research. Of course, we also cannot ignore the fact that the multinationals behind generative AI, such as Microsoft and Google, are looking to make a profit, not to help us in our mission as education professionals.

In this way, I hope to contribute to opening a space for debate in view of a "third path" between the neoliberal and the romantic approaches. The pragmatic and democratic conception I propose aims, in fact, to protect our profession from the excessive influence of all private, ideological and militant interests, as well as from a polarization about AI that risks harming the broad missions of higher education as a public good.

This conception recognizes the place of individual autonomy and the collective benefits of digital technology in all spheres of social life. And the same is true of AI, which occupies and will occupy essential functions in human development with not only important benefits, but also poorly known losses and risks, which we have a duty to identify as soon as possible to prevent them and minimize their harmful effects. Very soon, I will contribute to this research and prospective watch with the specialized teams of the International Observatory on the Societal Impacts of AI and Digital Technology (OBVIA). For now, I propose, as a conclusion, three main orientations for the smoother integration of AI in higher education.

Ongoing training on digital tools and AI. The vast majority of college teaching staff lack the essential knowledge and skills related to digital tools and AI to fully understand and embrace them. The ministère de l'Enseignement supérieur should invest in

[6] As defined in Quebec law or the *Global Convention on the Recognition of Qualifications concerning Higher Education* (UNESCO, 2019)

developing a continuing education program, designed by experts in AI and technopedagogy, that is well suited to the needs of the college system. A budgeted release time stipend program would allow professionals time and support to acquire a foundation that would then allow them to become self-sufficient in their disciplines—all of which are concerned.

Certified virtual assistants. As we illustrated with the collaborative approach, generative AI models are accessible to all organizations and offer novel opportunities to support faculty work. While these tools can provide some assistance, they are not expert software agents designed for teachers' tasks; they are often misused and sometimes ill-suited for higher education, and they raise a variety of security issues that cannot be detailed here. These tools also serve criticisable economic interests, notably the creation of new markets for multinationals via captive clienteles who are then encouraged or even forced to adopt their products according to harmful competitive dynamics.

We should therefore develop AI agents specialized in education, capable of assisting us in correction, supervision, feedback and pedagogical help. This would free up valuable time for rewarding and stimulating tasks, such as pedagogical innovation, personalized teaching, updating the content and knowledge of our courses, research and creation, etc. Whether developed by the state or by regulated companies, beneficial virtual assistants in higher education should, however, be subject to a certification process to ensure their effectiveness, safety and proven pedagogical benefits. In the meantime, departments and organizations in colleges and universities should remain vigilant and critical of the neoliberal rhetoric that asks us to adapt blissfully, in defiance of our professional autonomy and the educational purposes of our institutions in the service of society's common good.

The beneficial collaboration of human and artificial intelligences. As a human-machine interaction (HMI) researcher, it seems clear to me that the rapid arrival of AI agents in education will increasingly be perceived as a threat, giving fuel to the romantic perspective. Research has well documented the reactions of distrust and mistrust associated with the anthropomorphization of artificial agents with intellectual and language capabilities (Damiano & Dumouchel, 2018). Our psychosocial dispositions thus lead us to view robotic and software agents as "outsiders" and "threats" to the protected circle of our "humanity," specifically to our autonomy, dignity, and sense of competence. However, research also shows that the fear of being replaced or supervised by machines diminishes considerably when individuals learn to collaborate with humble and helpful assistants, designed to integrate seamlessly with human functioning.

To increase the chances of achieving an integrated collaboration of human and artificial intelligences, and to create a climate of trust and openness in the face of these novel and destabilizing technologies, it will be necessary to develop a strategy that takes into account this great challenge and undertaking. Above all, as prescribed by the *Montreal Declaration for a Responsible Development of Artificial Intelligence*, scaffolded in 2017-2018, it is imperative that principles such as autonomy, responsibility, well-being and inclusion be taken seriously in the integration of AI in human settings such as education. As far as I'm concerned, as education professionals, unions, and managers, we must demand the right to make real choices rooted in our core mission and professionalism. The power of generative AI will have both positive and negative consequences for most social systems, and together we have a responsibility to ensure that higher education is not simply a testing ground and profit-making source for multinational technology companies. ▬

References

Ahsan, K., S. Akbar, and B. Kam (2022). "Contract cheating in higher education: a systematic literature review and future research agenda," *Assessment & Evaluation in Higher Education*, vol. 47, n° 4, p. 523-539.

Bretag, T. and S. Mahmud (2009). "A model for determining student plagiarism: electronic detection and academic judgment," *Journal University Teaching & Learning Practice*, vol. 6, n° 1, p. 49-60.

Brynjolfsson, E. (2023). "The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence," *Dædalus*, Digital Economy Lab, Stanford University, January 12 [online].

Bubek, S. *et al*. (2023). "Sparks of Artificial General Intelligence: Early experiments with GPT-4."

Damiano, L. and P. Dumouchel (2018). "Anthropomorphism in Human-Robot Co-evolution," *Frontiers in Psychology*, vol. 9, p. 1-9.

Eaton, S. E. and J. Christensen Hughes (Eds.) (2022). *Academic integrity in Canada: An enduring and essential challenge*, Singapore, Springer.

Google Research (2023). "MusicLM: Generating Music From Text."

Martin, E. and S. Mussi (2023). *Bienvenue dans la machine. Enseigner à l'ère numérique*, Montréal, Écosociété.

OpenAI (2023). GPT-4 Technical Report.

UNESCO (2019). *Global convention on the recognition of qualifications concerning higher education*, Paris, France.

Surahman, E. and T.-H. Wang (2022). "Academic dishonesty and trustworthy assessment in online learning: A systematic literature review," *Journal of Computer Assisted Learning*, vol. 38, n° 6, pp. 1535-1553.

Wei, J. *et al*. (2022). "Emergent Abilities of Large Language Models," *Transactions on Machine Learning Research (TMLR)*, August.

Dave Anctil, PhD. teaches philosophy and artificial intelligence at Collège Jean-de-Brébeuf. He is also a researcher affiliated with the International Observatory on the Societal Impacts of AI and Digital Technology (OBVIA). Previously, he was a postdoctoral researcher and lecturer at the Université de Montréal and the Université du Québec à Montréal. His recent research and publications focus on the evolution and interactions of AI with humans, from the perspective of assessing psychosocial risks, benefits and transformations. He also works as an ethicist, popularizer and expert consultant for Canadian organizations and news media on ethical and psychosocial issues of AI and digital technology.

dave.anctil@brebeuf.qc.ca