The European Journal of International Law Vol. 34 no. 2

Editorial

Editorial: ChatGPT and Law Exams; On My Way In IV: 'Aren't You Exclusive?!' On the Pros and Cons of Writing Letters of Reference for Only One Candidate in an Academic Hiring Process; In This Issue; In This Issue – Reviews

ChatGPT and Law Exams

To suggest that AI is upending our world in a myriad of ways is by now a banality. To suggest that it poses a challenge to the very human condition, perhaps more so than previous technological revolutions, is, if not a banality at least a matter of extensive public discussion and debate. That there are no easy, consensus solutions has become self-evident. Here is but one much discussed example. Set aside the issue of deception. Imagine an AI that produces, say, a piece of music, or a painting, or poem. Openly and transparently. Acknowledging that it is in the 'school of' Mozart, or Titian or Szymborska, respectively. Imagine further that the 'machine' did a really good job. Would you, should you, enjoy it differently than you would if it were produced by a human, flesh and blood? The literature is conflicted, and I cannot even disentangle my own reactions to this. This, as mentioned, is but one tiny example. It will not be a short, or easy, process of civilizational and, perhaps, regulatory adjustment.

Be all that as it may, the advent of AI in general and one of its most accessible tools, ChatGPT, in particular, poses some daunting challenges in the world of higher education, not least when it comes to setting exams. Here we cannot set aside the issue of deception. In replying to an exam or writing a term paper, or Master's thesis etc., a student using ChatGPT without acknowledging such is simply cheating – their teachers and institution, their fellow students and, in a deeper sense, themselves.

The alarm is spreading: What are we to do when we set written exams, especially when these take the form of 'open book', and even more so when they are 'take home' (a practice very common in the US, less so in other jurisdictions)?

This is no longer a case of a hypothetical future. Colleagues from around the world are reporting increasing use, or suspicion of use, of ChatGPT by students. And since oftentimes the 'young' are ahead of the curve when it comes to the use of technology, the problem is likely to grow exponentially. (In an aside, I should mention that in our Journals we have already received articles about which we are sure they were written by GPT, and who knows if there are some we did not spot. More about this particular issue in a future Editorial).

In a recent faculty seminar at New York University Law School, we were shown examples of the potential of the current version of ChatGPT (3.5) to tackle even sophisticated exam questions. The results in some cases were impressive. Even when the AI reply was not perfect, deserving say an A, in several cases it would certainly yield a passing and even above average B mark. ChatGPT 4.0 is already available and is said to have enhanced capabilities, rendering the challenge even more acute. You (and your faculty) would be well advised to put it to the test with some real questions used in prior years' exams. You will almost certainly be surprised and even, perhaps, shocked.

As mentioned above, current discussions in different fora define the problem when it comes to exams, with obvious merit, as one of cheating and fairness. An exam, after all, is meant to test the abilities of the student and having ChatGPT answer the questions is not all that different from asking someone else, perhaps with more experience, to write one's answers. It is true that ChatGPT can also be used as a search engine – not the best of them. But since the lines to be drawn are pretty porous at this stage of the game, the undisclosed use of ChatGPT in replying to exams should be avoided and probably should be altogether prohibited.

When it comes to testing and grading student work, the ChatGPT challenge presents itself differently in relation to exams (whether in class or 'take home') and seminar papers. I will immediately state that the seminar paper challenge is far more daunting – ChatGPT is 'at its best' when it is asked to write an essay or at least a certain type of essay and answer open-ended questions. I will defer addressing the seminar paper challenge to another time. As mentioned, the challenge here is far more daunting given the strengths and weaknesses of the software. Here I will confine my reflections to exams *stricto sensu*.

We are in the early days of AI and higher education and it will take time for the Academy to adapt. But the problem of exams is already with us, and the following might be seen as no more than possible 'stop gap' solutions pending long-term adaptation.

The most common solutions discussed are those which would ban the usage of ChatGPT in exams and find ways of enforcing such.

In an earlier Editorial, 'Advice to young scholars VII: Taking exams seriously' (volume 33:1), I expressed my scepticism as regards the 20–30-minute oral exams as a mode of assessment, as practised in quite a few jurisdictions. At least regarding this issue (ChatGPT), my colleagues who were displeased with my critique of oral exams must be smiling. But, for better or worse, given my critique of oral exams, I cannot offer this as 'the' solution to the problem. One would be sacrificing too much, in my view. As indicated in that Editorial, a combination of written and oral examination (as we do with doctoral dissertations) would probably be ideal, but not practical in many situations and institutions.

So yes, I do think that as regards other forms of exams, one should prohibit the use of ChatGPT. Enforcement, however, is not as easy as it may appear. In most universities with which I am familiar, in in-class exams students are allowed to use their computers. Even without ChatGPT this poses a challenge to teachers or institutions who still believe (I do not) in 'closed book' exams. Still, we are familiar with

technologies that purport to prevent in-class exam students from accessing internet resources and even downloading resources on their computers. We also know that for many 'blocking technologies' there emerges rapidly a 'counter blocking technology' and it is never clear who is ahead of the game.

For online teaching, which is increasingly common (even post Covid), this challenge is even greater. Online courses and exams mean that by definition it is a 'take home' exam. No matter how sophisticated our online exam software is, how are we, if we take this prohibition approach, to ensure that the student is not using a secondary device on which they will happily ChatGPT (used as a verb) and then transfer its output to their principal computer?

One magic solution being discussed is to require, by regulation, platforms making ChatGPT type programs to have an automatic watermark on the products of such. It is a chimera. What is to stop the cheating student from producing a ChatGPT text (with the watermark) and just manually copying it to their exam or paper? (Cutting and pasting will not work as the watermark will remain).

I recently experienced the online challenge in person. I am currently enrolled, as a student, in an online master's degree. In taking exams we were required to write the exam by hand (!) and to have two devices (phone, iPad, etc.) with the cameras turned on, enabling the proctors to ensure that we were not using extraneous resources. One does not know whether to laugh or cry…. And if the proctoring is not human, another set of issues are likely to emerge as we know from the various shortcoming of face recognition and the like.

And yet, all this reminds me somewhat of the shrill cries of woe when cheap handheld electronic calculators were introduced in the 1970s and pupils began to bring them to school. Ban them, banish them, prohibit them was the initial reaction. It was a losing battle. The assumption today is that pupils even in grade school will make use of handheld calculators (present in every smart phone) and we simply adapted the way we teach and examine arithmetic and mathematics.

Be this as it may, it is an occasion to remind our students about integrity. Here is a formulation which resonated with me.

> If you cheat in this class, the best-case scenario for you is that you aren't caught, and that your [GPA] goes up a little…. [A]fter your first job, no one will ever ask you for your GPA again. To buy those temporary [advantages], you'll have practiced running roughshod over your conscience, making it easier to ignore that still small voice the next time. (L. L. Sargeant, 'Cheating with ChatGPT', *First Things*, 6 April 2023, https://www.firstthings.com/web-exclusives/2023/04/cheating-with-chatgpt).

Will it solve the problem? You might smile or grimace cynically, but this does not mean that this kind of message should not be heard in the classroom. We are educators, after all.

I want to suggest, tentatively, an altogether different approach to the ChatGPT exam problem. Like in medicine, where AI is used increasingly in diagnostic procedures, AI is most likely, in one form or another, to become part of legal practice. In fact, it is already there. If this is the case, we would be remiss if in our design of courses (and consequently exams) we stick our heads in the sand and pretend that AI does not exist. By

this I do not mean that we simply have to have, as already is the case, courses dealing with AI and the law, raising issues such as liability, e.g. when AI screws up and creates damage. I mean, instead, the integration of AI, including ChatGPT, into the regular teaching of all legal subjects.

So first some good news. At least in the current state of AI, including ChatGPT, it will not obviate the need for legal education and even, surprisingly perhaps, in the classical ways in which it is conducted now. To simplify just a bit, and again I borrow somewhat from the experience in medicine, one needs to 'know the law' in order to know how to use ChatGPT effectively: what questions to ask (what prompts to give is the lingo), what tasks to assign to AI, how to read its output intelligently and responsibly.

So no, you will not have to change radically your teaching. At least for now we, teachers, will not be out of a job.

This, however, does not mean that we could go on teaching in exactly the same way we do now, whether you belong to the lecturing guild or to the interactive so-called Socratic guild. You will, by whatever method, have to teach and train your students in the uses and abuses of AI and ChatGPT and similar programs which will no doubt emerge. Not all that different from the way we teach our students how to use LexisNexis, Westlaw and other 'analogical' online resources for legal research.

I want to illustrate this by one possible example and then move on to exams. Let's say you have finished teaching some area of positive law – say state responsibility (in international law) with the myriad provisions of the Draft Articles. A good teacher will want to satisfy themself that their students have not only understood the various provisions but that they have the capacity to analyse a complex problem of state responsibility, identify the issues it raises and choose correctly the relevant provisions of the Draft Articles and finally apply such to these issues with sophistication. So in one form or another, in good teaching, there would have to be a 'practicum' which follows the doctrinal segment of our teaching. It can be, for example, in the form of a hypothetical, or analysis of a complex case, in class or as homework. There is nothing radical or esoteric in that.

Enter ChatGPT. We might, for example, go through the hypothetical or the case, and ask the students how they would frame the ChatGPT inquiry to get the most effective responses. We might GPT (used as a verb) the problem or the issues ourselves and present to the students the result, and then analyse where it was helpful and where it was misleading. The students will thus experience both the potential and the risks of ChatGPT usage. I think my readers get the idea. I will add that this type of exercise will also hone the ability of the teacher to write sophisticated exams without GPT anxiety.

Here, then, are some thoughts on how one may approach exam writing in the GPT era. I do not want to pretend that I have found the holy grail, but if one accepts the view that in one form or another AI is here to stay and will be part of higher legal education, and in one form or another part of legal practice, this discussion must begin. My suggestions are intended to stimulate such discussion, solicit critique and produce further and perhaps more sophisticated approaches.

I am going to suggest now three, non-mutually exclusive ways of tackling the GPT problem in legal exams.

1. Take a slew of questions of different types (see my Editorial on exams) you have used in previous exams. GPT them yourself. You will soon enough discover the kind of questions the AI answers well or at least acceptably, and those where it does a very poor job. You will, thus, identify one kind of question you can set in a GPT anxious-free manner. As the software develops, you will need to repeat this exercise from time to time. When in your exam you have set a question that falls in this category, GPT it yourself as a reality check. You are most likely to discover, as I learnt also from our very own guru of law and technology, Thomas Streinz, that ChatGPT works best with open-ended essay-type exam questions, and worst with complex facts-type problems.
2. You might discover that there is a category of questions you like to give on which GPT does an OK job. Maybe not an A, but, say, a B- or B. Consider the following possibility. You set the question, you GPT it yourself and in your exam you set out the question and the GPT answer. The task of the students, on which they will be graded, would be to identify the various weaknesses or mistakes in the GPT answer. A variation on this theme, suggested in a very thoughtful Tweet by Dave Hoffmann, would be to ask the students to write an improved version of the ChatGPT product. One advantage of this approach would be that it may well replicate the type of usage made of GPT in law firms. This works for both doctrinal and non-doctrinal questions.
3. Assuming you have integrated GPT into your course design, and wish to test the students' proficiency, here is another possibility. Set the question (to repeat, it should not be of an essay type, such as 'To what extent does international law-making enjoy democratic legitimacy?'). The task for the students would be to break down the question to its components and formulate the most effective GPT prompts to be used. This is a reminder to our students (and to us) that without having studied law well in the traditional way, GPT is, if not useless, a defective tool.

I imagine that many would not wish to go so far down this route, but we should keep in mind that the use of AI in legal practice is likely to grow and that some of our students, regardless of our preaching, will make use of it. One way or another, one cannot escape the challenge.

JHHW

On My Way In IV: 'Aren't You Exclusive?!' On the Pros and Cons of Writing Letters of Reference for Only One Candidate in an Academic Hiring Process

At the time of writing – there is often a significant gap between the writing and publication of EJIL editorials – it feels like hiring season. Requests for letters in support of academic job applications pop up with the same speed as files to read for hiring panels. One question about academic practices has come up in discussions with both requesters of letters and letter writers: Should one be exclusive in letter writing, that is to say, should one write for only one candidate in any given application procedure?

As part of the EJIL On-My-Way-Out/On-My-Way-In series on academic practices, this Editorial reflects on some of the pros and cons of exclusive letter writing. It may seem a relatively minor question, but it can be a consequential one.

A Different Case: Writing for Admissions

Let us first distinguish writing for admission to masters and PhD programmes from academic job applications. Admissions are less of a zero-sum game than hiring processes, so there is less of an argument for exclusive writing. Indeed, there is a strong argument for non-exclusive writing: students often have only a few people who know them well enough academically to write a relevant letter of reference. Their teachers can therefore consider writing letters of reference as part of their duties: teach (and ideally inspire!), mark and write letters, lots of them. By contrast, colleagues in academic posts who are applying for a different job should have access to more people whom they can ask for an academic reference. But here the boundary begins to blur: when (former) PhD candidates or post-docs apply for their first academic position, does one write for only one of them if several apply for the same position?

Why Exclusivity?

One obvious and practical argument in favour of exclusivity is workload management. Writing individual letters of reference, tailored to a specific job, takes a lot of time, especially if one has to multiply that by several candidates. A more principled argument that is sometimes made is that it is not fair for the 'weaker candidate' if one also writes for a 'stronger candidate'. This argument resonates with the answer sometimes given in response to the slightly different question whether one should write at all if one cannot give one's full support: writing a letter that does not 'sell 100%' would not be fair vis-à-vis the person who asked for the letter. To answer these questions, we need to answer a prior question: To whom does one owe duties when writing a letter?

Balancing Duties

As Joseph Weiler observed in a previous Editorial on letters of reference, 'A balance needs to be struck between helping the candidate in his or her application purpose and an academic fiduciary duty owed to the admitting institutions in their selection procedures'.

From the perspective of a hiring institution, authors who write for several candidates can be tremendously helpful. Precisely because they know the differences between the applicants, they can and will often put different emphases – collegiality, scholarly breadth, scholarly depth, agenda setting, support for students, community service, originality, creativity, broadcasting power. Different people will come with different associations and experiences. When one and the same author writes very different letters for different people, the differences in the letters are clearly reflective of the differences in the applicants, rather than being attributable to letter-writing style.

While tremendously helpful to the hiring institution, writing for several applicants is not necessarily a breach of duty towards the applicant: candidates are different.

Thankfully, writing references is not a check-box exercise where referees fill in some standard template. It is a holistic reflection on the unique qualities of the particular candidate and the specific ways in which they would enrich the community they are seeking to join. Let the differences among the candidates appear through the implied comparison. Taking into account these different emphases, the hiring institution can then decide who would make the best fit.

That said, an issue does arise if one knows some of the letter-requesting applicants much better than others. The more one knows, the easier it is to elaborate. And in some academic cultures, the length and depth of letters are taken as indicators of support. A letter that extensively discusses publications, teaching and academic citizenship may thus be read as more supportive than half a page of comments on the CV and a reference to an encounter at a conference. This risk also exists if one was asked by only one candidate whose work one is not very familiar with, but it is exacerbated if one writes for more than one applicant because it will be apparent to the hiring panel that the brevity is not inherent in the author's writing style. It seems fair to be entirely open about this risk to the letter-requesting applicant and leave it to them to decide whether they still want you to write.

The situation gets particularly tricky if one deems one letter-requesting applicant stronger than the other requesters in almost all relevant aspects. But in a way, this dilemma is not much different from the situation in which somebody asks you for your support while your support is only lukewarm. The dilemma is only accentuated because a comparison between the letters makes it easier to gauge the different levels of support. In these situations, too, to put it in Weilerian terms, the truth may be the best lie: 'I can write for you, but given the job and the field (and in case one writes for others, too: "and the letters I am writing for other candidates"), I may not be able to give you enough support.' The applicant can then make an informed choice.

The greatest tension with duties vis-à-vis the letter-requesting applicant may arise if the author engages in explicit ranking: Y is my top candidate; Z no. 2; X no. 3. But it is even questionable whether such a ranking would be most helpful to the hiring institution. The ranking will be based on what the author values, not necessarily on what the hiring institution considers most important for the specific position. Compare it with peer review reports: unless editors have outsourced the decision-making to the reviewers, the most helpful reviews are not the conclusionary ones (accept/reject) but the ones that explain the evaluation, leaving it to the editors to weigh these factors to reach a decision in line with their editorial policy. Similarly, the whole range of considerations of the hiring institution are unlikely to be known to the letter writer, so one can leave the ranking to the institution.

Against Exclusivity

There are arguments against exclusive letter writing. First, unless one knows upfront exactly who will be in the race for which job, exclusivity could easily mean rewarding those applicants who ask first. Whilst in some contexts speed does deserve to be rewarded, it is not necessarily the case that the applicant who asks for a letter the moment the job ad appears is more worthy of a letter than the one who takes some time

to reflect. If the slower / more reflective applicant then asks the exclusive letter writer for a reference, it will be difficult for the writer to backtrack on the earlier commitment, even if the writer in fact thinks the latter would be a better match for the job. De facto, exclusive letter writing could then also risk privileging men over women, given that women tend to take the required qualifications more seriously than men and may need some more persuasion to apply.

Secondly, a principle of exclusive letter writing combined with a first-come-first-served approach may create question-begging silences: Why does this applicant not have a letter from, say, their former supervisor? Did they not want to write for them?

Thirdly, a principle of exclusive letter writing could give the wrong impression that letter writing is about championing a candidate by tying one's name to that candidate. In this understanding, the race seems to be one among the letter writers, rather than the candidates: Who gets their favourite in the post? The balance between the duty vis-à-vis the applicant and the institution has then been lost.

Finally, a risk of exclusive reference letter writing is that it could sustain … exclusivity. Many people on appointment panels look at the names on the letters: for better or for worse, big name and big institution are often taken to weigh more. If one is at such a big institution and chooses to write exclusively, one may find oneself constantly writing for one's own supervisees (to avoid the earlier mentioned silence question) or people in related well-networked institutions, rather than for scholars whose work one values and who have not passed through or worked at those institutions. Non-exclusive writing could do more to break with – or at least push against perpetuating – institutional path dependencies and academic hierarchies.

From a Binary to a More Communitarian Balance

Perhaps the balance that needs to be struck should not merely take into account the interests of one applicant and the hiring institution, but also those of many other applicants and the wider academic community. Exclusive letter writing risks reinforcing a view of academia that is dynastic and tri-lateral (candidate – letter writer – hiring institution) rather than communitarian.

None of this is to say that one should say 'yes' to every request for a reference. There remain plenty of reasons for saying 'no', including 'I barely know you or your work', or leaving the choice to the applicant after saying 'I am also writing for someone else for whom I can provide more evidence on all the selection criteria'. The only argument that would fall by the wayside is that of exclusivity as a principle. Let the ink flow in more directions.

SMHN

In This Issue

EJIL's year-long symposium 'Re-Theorizing International Organizations Law' continues in this issue with two articles that put the spotlight on thinkers of international organizations law beyond the usual suspects. *Kehinde Olaoye* introduces Samuel K.B.

Asante's academic writings and experience as an international civil servant in a now-defunct UN unit specializing in transnational corporations. Olaoye takes Asante's intellectual and professional trajectory as an occasion to reflect on the place of Third World approaches in international organizations law. In the next contribution to the symposium, *Francisco-José Quintana* examines the legacy of Jorge Castañeda as a semi-peripheral jurist whose work signals the many ways international organizations can harm, empower or elude the control of smaller states. Castañeda's legal and political outlook, Quintana argues, remains ever relevant to debates on universalism, regionalism and power asymmetries in international organizations.

Roaming Charges in this issue pictures a moment on a Sunday afternoon in Washington Square, New York: two individuals immersed in their individual yet common realities.

In this issue's Articles section, *Eliav Lieblich* asks whether states can use force to recover territory during prolonged occupation. Lieblich disentangles the implications of the permissive and restrictive approaches and the normative commitments required of each camp. He makes the case for an individualist sensibility that prioritizes the avoidance of human suffering over territorial integrity, arguing that the crux of such 'wars of recovery' lies in the tension between statism and individualism.

Exploring another polarity between states and natural or legal persons, *Yohei Okada* analyses how the distinction between *acta jure imperii* and *acta jure gestionis* influences the attribution of the conduct of parastatal entities to a state. For Okada, Article 5 ARSIWA has been read too narrowly as precluding the attribution of commercial acts of parastatal entities to the state. The article advances a different reading of that provision and describes the scenarios in which *acta jure gestionis* may be attributable.

In the next article, *Filip Batselé* unearths archival material from the 1950s and 1960s that sheds light on the lobbying for private foreign investors carried out by a little-known association: the Association for the Promotion and Protection of Private Foreign Investments. Batselé recounts how that group of lawyers and businessmen came together, with varying degrees of success, around the goal of influencing the laws of foreign investment protection.

The following article studies another attempt to influence law-making. *Bruno Biazatti* analyses the replication of the wording of existing treaties as a drafting technique, taking the International Law Commission's Draft Articles on crimes against humanity as a case study. For Biazatti, the adoption of familiar formulae found in other legal instruments reflects a political choice of the Commission; the jury is out on whether states will back the ILC's 'copy and pasting'.

Closing the Articles section, *Itamar Mann* focuses on the struggle of Palestinian refugees. The exclusion clause in the Refugee Convention, Mann warns, is unfit for the unique situation of Palestinian refugees registered with the UN Relief and Works Agency. Placing the question in the broader context of the Palestinian fight for self-determination, the article invites us to reconsider the political foundations of refugee law.

The Last Page in this issue presents a poem by a Portuguese poet, *Florbela Espanca* (1894–1930), beautifully translated by *Kay Cosgrove*. A feminist spirit

living in turbulent times, 'To Be a Poet', Espanca wrote, 'is to condense the world into a single cry'.

ALB

In This Issue – Reviews

In this issue, we have something for everyone to inform your reading in the six reviews of recent books.

We begin with two books that address the law of the sea, but they do so from very different angles. *Douglas Guilfoyle* reviews Ian Urbina's 'vivid and often confronting' book, *The Outlaw Ocean*, a book which seems possible to read through different lenses, including those of the activist scholar and practising lawyer. While observing that of course law is 'complicit in creating structures that facilitate exploitation and abuse', Guilfoyle notes that this book is read not 'for legal analysis but to bear witness to a reality concealed from those who live their lives ashore', a reality which includes inhuman hours, dangerous conditions, unhealthy environments, violence and exploitation. In comparison, the second law of the sea book takes us back to the more predictable shores of the specialized genre of legal academic writing. Lorenzo Palestini's *La protection des intérêts juridiques de l'État tiers dans le procès de délimitation maritime* seeks to doctrinally examine in detail the place of third states in maritime delimitation processes. *Massimo Lando* reviews this French-language text, noting its usefulness for practitioners as a systematization of an unpredictable jurisprudence.

Two books focusing on the history of international law make up the middle part of this review section. First is *Jochen von Bernstorff*'s review of Ntina Tzouvala's 'original contribution', *Capitalism as Civilization.* The book, he notes, 'provides us with a critical re-description of the concept of civilization in international legal discourse across time informed by Marxist, postcolonial and feminist theories'. *Miloš Vec* reviews a volume edited by Marcus M. Payk and Kim Christian Priemel, *Crafting the International Order,* which carefully sketches the role of practitioners in the making of international law. The review faithfully sketches in turn each of the chapters, but also notes what is missing from the volume, such as the stories of female practitioners.

Two further reviews complete the section. *Dana Schmalz* finds Liv Feijen's account of *The Evolution of Humanitarian Protection in European Law and Practice* a 'highly informative read', but would have preferred a fuller engagement with the complex notion of 'the humanitarian', which remains rather too 'vague' for her taste. Richard Gaskins' work on *The Congo Trials in the International Criminal Court* is the final book under review. *Raphael Oidtmann* situates both the book and the trials, applauding the consideration Gaskins gives to Ituri, the locus of many of the crimes before the ICC, which is often neglected in scholarship on the trials.

Happy reading!

GCL and CJT