Received: 23 January 2024 | Revised: 2 May 2024 | Accepted: 4 May 2024

DOI: 10.1111/1467-8500.12647

WILEY

RESEARCH NOTE

ChatGPT in public policy teaching and assessment: An examination of opportunities and challenges

Daniel Casey

School of Politics and International Relations, Australian National University, Canberra, Australia

Correspondence
Daniel Casey, School of Politics and International Relations, Australian National University, Canberra, Australia.
Email: daniel.casey@anu.edu.au

Abstract

This paper presents the findings of an innovative assessment task that required students to use ChatGPT for drafting a policy brief to an Australian Government minister. The study explores how future public policy students perceive ChatGPT's role in both public policy and teaching and assessment. Through self-reflective essays and focus group discussions, the research looks at the limitations of ChatGPT that the students identified, demonstrating it struggles to produce analytically sound, politically responsive, and nuanced policy recommendations. The findings align with the "technoscepticism" theoretical frame, indicating concerns that artificial intelligence (AI) tools could undermine good policy analysis processes. The students supported greater use of ChatGPT in the classroom, to increase ChatGPT-literacy, help students learn to engage ethically and appropriately with AI tools, and better develop evaluative judgement skills.

The paper contributes insights into the intersection of ChatGPT, teaching and assessment, and public policy and seeks to prompt further exploration and discussion

on the implications of integrating ChatGPT into both public policy and its education and assessment.

KEYWORDS
AI literacy, assessment design, ChatGPT, public policy education, technoscepticism

Points for practitioners

- Future public service graduates are highly sceptical about the value of ChatGPT for developing policy.
- They are concerned about the ethical implications, the lack of transparency, and the impact it may have on marginalised communities.

1 | INTRODUCTION

What happens when 25 students are tasked to use ChatGPT to write a brief to an Australian Government minister on a topic area of their choosing? Does it produce useful, well written briefs, with thoughtful, nuanced, politically responsive recommendations? Or does it produce recommendations to re-establish the long-defunct Australian domestic car manufacturing industry (Participant 1)[1] and falsely accuse a senior public servant of serious conflicts-of-interest (Participant 5)?

The launch of ChatGPT in late 2022 lead to an explosion of interest in generative artificial intelligence (AI) and the possibilities it creates. That was quickly followed by concern across schools and universities about the risks to academic integrity. One survey in early 2023 indicated that one-third of students were using ChatGPT in some way in their essay writing (Sullivan et al., 2023). Within the Australian government, the Department of Home Affairs has been trialing the use of large language models (LLMs) (Evans, 2023), but (as at April 2023) "there is no current [whole of Government] policy on the use of generative AI technologies, such as ChatGPT" (Commonwealth, 2023). For scholars of public administration and public policy, ChatGPT presents multiple overlapping challenges, firstly relating to the use of ChatGPT as academics, researchers and teachers, and secondly, normative and empirical questions about the use of ChatGPT inside governments. For a discipline committed to a close link between our research, teaching and practice (St. Denny & Zittoun, 2024), these issues are intricately linked, and our Scholarship of Learning and Teaching (SoTL) will need to urgently adapt.

However, there is a lack of focus on SoTL in public administration more broadly (McDonald et al., 2024) and specifically a lack of empirical research on how to adapt our teaching methods to incorporate ChatGPT. This paper sets out the findings of an innovative assessment task undertaken as part of the capstone public policy subject at the Australian National University, which required students to use ChatGPT to write a policy brief to an Australian Government minister, and then write a personal reflection on the process. The assessment was designed to explore how future public policy and public administration professionals view the role of ChatGPT in their future workplaces, as well as to consider how to best use ChatGPT as part of teaching and assessment. Using these essays and a subsequent focus group discussion, I find that while students were

initially positive about ChatGPT, over the course of the assessment they realised the significant limitations of these tools, and the significant dangers that overreliance on them can create. This aligns with the "technoscepticism" theoretical frame that Newman and Mintrom (2023) develop, which suggest that AI tools undermine good policy analysis processes, because of the lack of human oversight and transparency. From a pedagogical perspective, I find that using ChatGPT in the classroom helped students develop specific AI-literacy skills, as well as general evaluative judgement (Tai et al., 2018) and critical thinking skills.

While this is a small, single university study, consistent with the role of research notes, this paper seeks to raise new and stimulating research questions and teaching approaches, which I hope will inspire further exploration of this teaching and assessment approach in the future.

2 | BACKGROUND AND RESEARCH QUESTIONS

This paper sits at the intersection of different literatures, including SoTL, theories of assessment design and graduate skills; the impact of ChatGPT on pedagogy and assessment; and normative questions about the use of ChatGPT inside government. I first examine the literature related to the impact of ChatGPT on teaching and assessment across disciplines, before moving onto the literature on SoTL in public policy, and ChatGPT and policy making.

Universities are struggling to work out how to deal with AI within the teaching and assessment environment (Cotton et al., 2023). The focus in mainstream media has been on policing ChatGPT usage, with a focus on issues associated with academic integrity, and preventing students using ChatGPT (e.g., a return to pen-and-paper exams). This framing of ChatGPT as a cheating tool, rather than a potential positive, may actually become a self-fulfilling prophecy (Sullivan et al., 2023). In response, universities have revised their academic integrity requirements. Many universities are using the *Turnitin* AI detector, while other universities have opted-out of this tool (Bates & Berringer, 2023). While *Turnitin* suggests that they have a false positive rate of less than 1%, and makes their determination with "98% confidence" (Turnitin, 2023), the advice provided by universities is mixed, with ANU emphasizing that it should be used "with caution" and is not yet a "reliable indicator of academic misconduct" (ANU, 2023).

While ChatGPT has and will continue to drive changes to assessment, it is up to academics to make sure that these changes are for the better, rather than simply returning to paper-based invigilated exams. The Australian regulator, the Tertiary Education Quality and Standards Agency (TEQSA), recently released a discussion paper on assessment in the age of AI (Lodge et al., 2023), which provides a range of propositions to reform assessment. These include, inter alia:

> ***Propositions***
>
> … *Assessment should encourage students to critically analyse AI's role in, and value for, work and study, aligned with disciplinary or professional values… [and] where learners critically engage with the use of AI, demonstrate judgement in how to best use AI and reflect on the learning process*. (Lodge et al., 2023, p. 4)

This proposition is, in large part, derived from existing theoretical approaches to assessment from the broader SoTL literature, including "authentic assessment", "constructivist learning" (Bada & Olusegun, 2015) and "evaluative judgement." I focus on evaluative judgement, which is the ability

to "judge the quality of one's own and others' work" (Tai et al., 2018). This skill is regarded as a crucial foundation for lifelong learning and improvement, as it enables individuals to operate independently. This is likely to become increasingly important in an AI-dominated world—where access to information itself is rapid and easy, the valuable skills will cease to be raw knowledge or information retrieval, but rather assessment and critical engagement with that information. This means that, across disciplines, teachers need to amend assessment tasks to reflect this skill (Bearman et al., 2020) and to prepare students for a "digital world… assessment needs to embrace" these new technologies (Dawson, 2020, p. 38).

Searching across disciplines, there are a few articles that report either on students' perspective of ChatGPT—Firat (2023), for example, looks at graduate student perspectives across a range of disciplines, with a focus on Turkey. Elkhodr et al. (2023) focuses specifically on ICT students, reporting on a project allowing and encouraging students to use ChatGPT as a tool in their assessment. However, there do not appear to be articles reporting on assessment tasks that require an essay to be entirely written by ChatGPT in any discipline. These early articles suggest a mixed perspective, with some suggesting significant benefits, while others highlight the risks. All, however, focus on the need to improve digital literacy skills of both students and academics. This leads to my first research question:

RQ1: What are students' perspectives on using ChatGPT in teaching and assessment in public policy?

The second issue is the impact of ChatGPT on public policy development. We are already in an age of "crises of expertise in liberal democracies," (Head, 2023, p. 1) driven, in part, by reduced policy capacity within the public service, and the contestability of advice from outside the public service. At the core of the quality policy advisory systems is substantive issue expertise (Migone & Howlett, 2023), which may be threatened by a greater reliance on AI. Others, however, are more positive,suggesting that ChatGPT can unleash the "augmentation of human intelligence" with AI (Dwivedi et al., 2023, p.55).

All tools (a hammer, an abacus or a calculator) are designed to improve human performance—"when humans use these tools… the human's cognitive ability is augmented" (Fulbright & Morrison, 2024, p. 1). Unlike calculators or hammers, however, there is a fear AI tools like ChatGPT could replace human intelligence, rather than augmenting it (Dwivedi et al., 2023). However, like any other tool, its capacity to improve both learning and outcomes depends on how they are used—working out in what circumstances it should augment human labour and discretion, and where it can safely supplant humans (Ahn & Chen, 2022).

With some high-profile apparent failures of automated decision-making (Casey & Maley, 2024 Newman & Mintrom, 2023; Whiteford, 2021), a cautious approach to the use of AI inside government has been adopted. The Australian Government is working through these issues and has sought feedback on the use of AI in the public sector (Department of Industry, Science and Resources, 2023). In response, many submitters argued that "public sector decision-making is held to higher standards" (Weatherall et al., 2023, p. 32) and that "the public sector has a greater responsibility to lead and to ensure that AI does not have a negative impact on society" particularly in relation to marginalized groups, such as Indigenous Australians (Gang Li et al., 2023, p. 7). Submitters called for a greater level of transparency and oversight of AI use inside government (Ombudsman, 2023) and to prioritize AI literacy and training within government (Falstein, 2023; Marsden et al., 2023). These submissions also drew attention to the clash between AI; "evidence-based"/"evidence-informed" policymaking; and the ethical practice of policy analysis.

The use of AI tools, such as ChatGPT, in policy-making "will upend the previous discourse on policy analysis and evidence-based policy" (Newman & Mintrom, 2023). As ChatGPT allows for the automation of more and more aspects of policy analysis, how does that intersect with public service ethical requirements, such as impartiality and accountability? While ChatGPT allows for analysis of big data faster than ever before, potentially allowing for more robust evidence-informed advice, the algorithmic black-box prevents the level of transparency that is usually expected of public policy (Newman & Mintrom, 2023). Analysing the discourse around AI and ChatGPT, Newman and Mintrom (2023) identified eight policy "frames," which demonstrate how the same issue can be understood differently. The two most relevant for this exercise are as follows:

Frame 1: Faith in Rationality… In short: artificial intelligence represents a technological advance in evidence-based policy making. These technologies can provide greater quantities of policy-relevant information than human policy analysts could, and much more quickly.

Frame 2: Technoscepticism…In short: artificial intelligence technologies undermine the quality of knowledge useful to making policy decisions, because the information cannot be independently verified (Newman & Mintrom, 2023, p. 1846).

However, there is a lack of knowledge of which frames are adopted and predominate in academia, within government, and within future practitioners. Similarly, while theorizing and hypothesizing has occurred about how ChatGPT can improve policy development (Cantens, 2024; Dwivedi et al., 2023; Huang & Huang, 2023), there is an absence of empirical research in this space.

If we are committed to embedding our research, and real-world challenges into our teaching, we need to understand the direction that governments are taking with the use of AI in policy development and embrace the challenge of incorporating this into our teaching. McDonald et al. (2024, p. 20) suggest that "artificial intelligence is becoming a prime tool in public servants' jobs," which means it is incumbent on academics and our institutions to equip our students accordingly. This includes helping them understand what AI can and cannot do, and what humans will still be expected to do (Michels, 2023). This leads to my second research question:

RQ2: What are students' perspectives on the impact of ChatGPT on the development of public policy?

3 | METHODOLOGY AND DATA

Applied Policy Project (POLS3041), is the capstone course for the Bachelor of Public Policy degree at the Australian National University. In most years, between 15 and 25 students take the course. It is largely self-directed, requiring students to choose a policy area/problem and undertake detailed research on the issue and prepare a range of different types of policy papers on their chosen topic. In previous years, the three assessment items were an "issues brief," and "options paper" and a "decision brief." In 2023, the final piece of assessment was changed, to require students to use ChatGPT[2] to draft the "decision brief." Consistent with the Australian Government's current requirements for policy development and Cabinet submissions, the brief was required to include a First Australians Impact Assessment statement (NIAA, 2023a).

Students then submitted a self-reflective essay, exploring the process of working with ChatGPT, including their thoughts on the use of ChatGPT to develop policy and write policy documents, and reflect on the process as a piece of assessment—What did they learn from this? Did they develop useful skills? There is significant evidence of the benefit of incorporating reflective essays into assessment practices (Allan & Driscoll, 2014). The act of self-reflection helps deepen and embed the learnings, helping students to make explicit their internal process. By "rethinking" their past actions, it provides a basis for critical thinking. The essays also gave me a window into the students' learning process (Allan & Driscoll, 2014), which also improves my teaching in subsequent years.

Prior to the assessment, I dedicated a 2-hour workshop to introduce ChatGPT, exploring how it is already being used within government, including a guest lecture by an Australian Senator who has taken a significant interest in the use of ChatGPT within the Australian Government.[3] I also provided a series of exercises for students to work through, in groups, to explore how ChatGPT worked.[4]

While all students undertook this assessment, seven of 24 students (30%) opted into this research.[5] These students were broadly representative of the entire class, with similar marks for this assessment (mean of 75 for the participant group, versus 74 for the entire class) and the gender balance was also representative (the participants were 70% female, the same as the entire class). Only these essays and their ChatGPT transcripts were used in this research. One focus group was also conducted.[6] Given this is an emerging field of research, an inductive coding approach was adopted for the essays, transcripts and focus group transcript (Chandra & Shang, 2019) to identify themes. ChatGPT was also used to identify themes.

4 | FINDINGS

Before turning to the main findings, I first look at how Turnitin handled these ChatGPT produced briefs. Despite Turnitin's accuracy claims, for the ChatGPT produced briefs, Turnitin's AI reports varied wildly. While approximately half of the essays resulted in an AI detection of 100% (as should be the case), around 25% produced AI detection of 50% or less, including three essays (13%) where the AI detection was below 10%. This aligns with extensive experimental evidence by Foster (2023) who managed to engineer prompts to systematically fool Turnitin's AI detection tool, including demonstrating that ChatGPT "knew" how to change its own writing to make it seem less like AI.

While students felt quite positive going into the exercise, with one noting how "intuitive" ChatGPT was (Participant 3), the overwhelming feeling at the end was disappointment that ChatGPT could not produce anything that would have received good grades, or that they would have been prepared to present to the minister. All students identified that ChatGPT produced something that looked good, that read well, but was shallow and lacking in any sort of analysis. Within the self-reflective essays and focus group discussion, six key themes came out (Table 1).

4.1 | Findings on research question 1: What are students' perspectives on using ChatGPT in teaching and assessment in public policy?

Firstly, during the focus group discussion, the participants talked about how the exercise helped them critique and evaluate another's work. They reflected that across the semester, they had

TABLE 1 Key themes emerging from the research.

Developing skills in assessing and critiquing others' work (research question 1)
Importance of learning how to best engage with ChatGPT and craft useful prompts/questions (research question 1)
Recognition that ChatGPT has some benefits for public servants and the policy process (research question 2)
Struggled to address issues associated with Indigenous Australians (research question 2)
Factual errors and hallucinations (research question 2)
Lack of political responsiveness and nuance (research question 2)

become semi-experts in their topic, so when they were confronted with an essay produced by ChatGPT, they were better able to identify the problems with it. I agreed with these evaluations, and none of the ChatGPT-produced briefs would have scored higher than a low credit. This disappointment ("I was not happy" [Participant 5]; "far from sufficient quality" [Essay 7]; "not be able to produce a reputable piece of work" [Essay 2]) reflects the development of their evaluative judgement.

The second theme that most students identified was the need to learn how to craft prompts and engage with this new tool, with one saying "[i]t is possible that with more practice… a higher quality brief could be generated" (Essay 7) and another noted that she used existing resources "to create prompts that are effective" (Essay 6). In the focus group, participants confirmed that the exercise helped them understand how to engage with ChatGPT and how to ask the right questions to get the answer that they wanted (Participants 2, 3, 5 and 7). One participant reflected that they spent a lot of time "trying to figure out how to get it to actually… answer the prompt" (Participant 5). Another compared it to the introduction of other "groundbreaking technology"—"it feels like somebody's introduced the calculator to us. And we're trying, we're learning what the limitations are of the calculator again" (Participant 7).

The transcripts demonstrate a wide variation in students' understanding of how to craft effective prompts. For example, some students failed to specify which minister (or which level of government) the brief was aimed at (Transcripts 1 and 3), or who the audience was (Transcript 4). While others provided prompts that more closely reflect what is considered good practice. For example, some students provided significant background information and context, such as the complete assessment tasking document and guidance on how to write a brief to a minister (Transcript 7), that they were an Australian university student (Transcript 3) or policy detail (Transcript 2). Most students then iterated with changes to structure and emphasis (Transcripts 1, 3, 5, 6, 7), while others did not iterate or provided very limited instructions (Transcript 4). One student asked ChatGPT how to improve their prompts "Do you have any constructive feedback for me?" (Transcript 3).

Overall, in response to this first research question I found the exercise led to general negativity and "pessimistic viewpoint on ChatGPT" (Participant 5). Nevertheless, there was broad support for the exercise itself, in part because it aligned with the guiding principles set out by TEQSA (Lodge et al., 2023), because it helped equip students to engage with AI, analysing AI's role in our discipline and consider the ethics and risk associated with its use. One participant said:

> I thought it was a really, really valuable exercise, like I'm kind of really grateful… And I thought that was, from a pedagogical perspective, was really valuable. I liked it (Participant 7).

4.2 | Findings on research question 2: What are students' perspectives on the impact of ChatGPT on the development of public policy?

Moving to the second research question, and issues specifically associated with public policy, all students recognised that ChatGPT has some benefits for public servants and the policy process by augmenting human intelligence, reducing cognitive load and saving time; it "can 'complement' the work of policy makers" (Essay 7), but it should "only be used under close supervision" (Essay 5). The remaining findings drew out the risks associated with ChatGPT, highlighting the importance of skilled operators, who understand both the substantive policy field and the limitations of ChatGPT.

ChatGPT struggled with the "First Australians Impact Assessment" section of the brief, and this was an area where differently constructed prompts resulted in very different results, but participants remained unsatisfied with ChatGPT's output (Essays 1, 4 and 6). Participants expressed concern that because the views of marginalized groups (including Indigenous Australians) are likely to be underrepresented in the training data, the output is likely to poorly represent their perspectives (Essay 2), while another raised a concern that ChatGPT is unlikely to meet requirements for data sovereignty for First Nations' knowledge (NIAA, 2023b) (Essay 7).

Next, and consistent with the existing literature, all students identified issues with misrepresentation of facts and data. One student specifically requested that ChatGPT add statistics, but then when the student queried the accuracy of the statistics, ChatGPT said that the statistics were "simulated or fictional statement(s) created for the purpose of the policy brief," and were "not based on any specific or actual meta-analysis or research" (Essay 3). During the focus group, another participant revealed that ChatGPT named and accused a senior Australian public servant[7] of serious conflicts-of-interest; however, the participant could not find any evidence of this (Participant 5), and it appears to be another example of an "artificial hallucination," which could have serious adverse consequences for the named individual, if people unquestioningly trust ChatGPT's output.

While many of the comments and criticisms about ChatGPT's products could apply to most sectors, some students identified that the lack of clarity and transparency in ChatGPT's processes is a particular issue for public policy and public administration—"[w]ithout a clear understanding of how information is generated in the system, transparency issues arise about what evidence the information is based on" (Essay 2). The students also identified that this risks "enhance[ing] and replicat[ing] pre-existing structural disparities in society" (Essay 2). When this was discussed in the focus group, "AI-informed policy" was contrasted with "evidence-informed policy" (Participant 3).

Similarly, many students commented that the briefs produced by ChatGPT were not responsive to the political environment or recognise existing Australian Government policies (Essay 2) or institutions/governance arrangements (Essay 5) in the policy area. One student (Essay 3) explicitly asked ChatGPT to rewrite the brief through a "conservative" or "progressive" lens—but found that ChatGPT only changed adjectives, adding the word "progressive" and "progressivism" 27 times in a 600-word brief, but did not make any substantive amendments. Similarly, when asked to rewrite the brief "from a conservative political perspective," it removed the word "progressive" and replaced it with the word "conservative" 18 times, but again without any substantive amendments.

Overall, in response to the second research question, *What are students' perspectives on the impact of ChatGPT on the development of public policy?*, I found that participants' perspectives

broadly aligned with the concerns expressed in the Australian Government's consultation process, including the importance of transparency, and disadvantaging already marginalized groups. These views fit with the "technoscepticism" frame of Newman and Mintrom (2023). Participants saw ChatGPT as undermining the benefits of evidence-based policy, introducing significant risks to the policy process.

5 | DISCUSSION AND CONCLUSIONS

In this paper, I report on an innovative assessment undertaken as part of the capstone public policy subject at the Australian National University, which required students to use ChatGPT to write a policy brief to an Australian Government minister, and then write a personal reflection on the process. The aims were to expose students to ChatGPT, get them to reflect on its strengths, weaknesses and possible role in both public policy and teaching/assessment. With the exponential take-up of ChatGPT, it is becoming increasingly important for academics to consider the normative and empirical questions around ChatGPT's usage in both government and teaching. This assessment challenged future practitioners to consider these questions themselves.

This task was deliberately unrealistic, requiring students to exclusively use ChatGPT to produce the brief, rather than using ChatGPT as a tool, partner, or team member (Dwivedi et al., 2023). It is far more likely that ChatGPT will be used as an example of "human cognitive augmentation" (Fulbright & Morrison, 2024), and thus the challenge will become how to best work with this new team member, understanding their strengths and weaknesses. However, by getting students to push ChatGPT to produce the brief by itself, it forced students to explore the limits of the technology (as it currently stands), which will help the students understand how to augment their own capabilities with this new technology.

St. Denny and Zittoun (2024) emphasised the importance of maintaining the link between our public policy research and our training of future policy practitioners. This makes it vital that our teaching methods adequately prepare our students for policy workplaces of the future. Unfortunately, we have "only begun to grapple with the pedagogical aspects… of artificial intelligence" (Bakir et al., 2024, p. 286). This research contributes to this space by exploring the role of ChatGPT in our teaching and perspectives of its role in future policy workplaces.

While a small sample, the findings add to the literature on the use of ChatGPT in both teaching and assessment, and in public policy making. From a pedagogical perspective, I provide a practical suggestion on ways to incorporate ChatGPT into the classroom. The findings also support further research on how to best incorporate ChatGPT-literacy, including prompt engineering, into the broader curriculum and the public policy/public administration curriculum in particular.

I agree with Illingworth (2023) that ChatGPT provides a chance to reconsider our broader approach to designing assessment, and that our challenge is to design assessments that are "authentic," meaningful, useful and relevant. This is likely to mean focusing on higher order critical thinking and evaluative judgement skills, which are becoming more important as AI-generated content challenges us to prevent the dissemination of mis- and disinformation. The process of critiquing a ChatGPT-produced essay has similar benefits to peer-review and peer-feedback, which is already well established in the literature (Tai et al., 2018). If others consider using this type of assessment in the future, I would suggest focusing more explicitly on improving *students'* evaluative judgement, asking them how the work "meets or does not meet agreed standards and criteria" (Tai et al., 2018). One participant suggested that this could include requiring students to edit/fix the ChatGPT produced brief, in track changes (Participant 5).

From a public policy and public administration perspective, the participants were pessimistic about the potential for ChatGPT to make significant contributions to policy development soon. This stands in stark contrast with the existing theoretical literature (Cantens, 2024; Dwivedi et al., 2023; Huang & Huang, 2023) that suggests a broad range of possibilities for ChatGPT in policy analysis and development. While they might not be quite the luddites of old, there was a healthy scepticism about the challenges of using ChatGPT in general, and the specific challenges within the public sector, because of the importance of transparency and accountability. However, further research with current policy practitioners is also warranted to see if they are similarly techno-sceptical.

I recognise that this research has a range of limitations. It was conducted with only a small group of students, at a single university, and students only used the free ChatGPT3.5, rather than the subscription-only ChatGPT4, which is likely to have produced better results. For academics thinking about similar assessment tasks, equity issues need to be considered—while some universities now provide subscriptions to ChatGPT4(HKU, 2024) or Microsoft co-pilot (ANU Centre for Learning and Teaching, 2024), many universities may not provide these services yet. The rapidly evolving abilities of generative AI means that the findings and conclusions here are, to a large extent, a moment-in-time. As generative AI improves, and students improve their prompt-engineering, it is only a matter of time before generative AI is able to overcome some of the issues identified here.

Given the fast-moving nature of AI development, there are benefits in spreading early findings, as a way of developing more extensive research agenda. Consistent with the objectives of research notes, this research should raise new and innovative approaches to both research and teaching, and I hope it will inspire further normative, empirical, and theoretical projects in the future.

ACKNOWLEDGEMENTS

I would like to thank Zoe Robinson, who encouraged me to run this assessment. An earlier version of this paper was presented at the 2023 Australian Political Studies Association Conference and I would like to thank Diana Perche for her feedback on that paper. I would also like to thank Phillip Dawson and Kate Elkins for their suggestions. I also thank the leadership of both the School of Politics and International Relations and the College of Arts and Social Sciences at the Australian National University who supported this project. ChatGPT was used to assist drafting the title and abstract, as well as in identifying common themes in participants' essays. All the writing was done by the author.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Daniel Casey https://orcid.org/0000-0003-2115-4431

ENDNOTES

[1] References to "participant" is where the source was the focus group. "Essay" is a reference to their self-reflective essay, and "transcript" is a reference to the full ChatGPT/Bard transcript they were required to submit. The same numbering system was applied across these three sources.

[2] Students could also use Google Bard. For convenience's sake, I refer to ChatGPT throughout.

[3] Senator David Shoebridge, an Australian Greens Senator from New South Wales.

[4] These exercises are included in the Appendix.

[5] The ethical aspects of this research were approved by the Australian National University (2023/270 refers). Ethical considerations are particularly important when using students as research subjects. Students sent consent forms directly to a different academic, who kept them until grades were released. That provided assurance to students that their grades weren't impacted by their decision to participate and addressed the power imbalance in research involving students.

[6] Consistent with the ethics approval, the focus group occurred after results were released, to minimise concerns around power imbalances between the researcher and students. Additional details on the focus group are included in the Appendix.

[7] In the interests of protecting the falsely accused individual, the Participant 5 did not share the name with me.

REFERENCES

Ahn, M. J., & Chen, Y.-C. (2022). Digital transformation toward AI-augmented public administration: The perception of government employees and the willingness to use AI in government. *Government Information Quarterly*, *39*(2), 101664.

Allan, E. G., & Driscoll, D. L. (2014). The three-fold benefit of reflective writing: Improving program assessment, student learning, and faculty professional development. *Assessing Writing*, *21*, 37–55. https://doi.org/10.1016/j.asw.2014.03.001

ANU. (2023). *AI writing detection now available in Turnitin*. https://telt.weblogs.anu.edu.au/ai-writing-detection-now-available-in-turnitin/

ANU Centre for Learning and Teaching. (2024). *AI quick-start guide for start of semester*. https://teaching.weblogs.anu.edu.au/resources/ai-quick-start-guide/

Bada, S. O., & Olusegun, S. (2015). Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education*, *5*(6), 66–70.

Bakir, C., Singh Bali, A., Howlett, M., Lewis, J. M., & Schmidt, S. (2024). Teaching policy design: Themes, topics and techniques. In E. St. Denny & P. Zittoun (Eds.), *Handbook of teaching public policy* (pp. 278–292). Edward Elgar Publishing. https://doi.org/10.4337/9781800378117

Bates, S., & Berringer, H. (2023). *A message from LT HUB leadership re. TURNITIN*. The University of British Columbia. https://ctl.ok.ubc.ca/2023/04/03/a-message-from-lt-hub-leadership-re-turnitin/

Bearman, M., Boud, D., & Ajjawi, R. (2020). New directions for assessment in a digital world. In *Re-imagining university assessment in a digital world* (pp. 7–18). Springer.

Cantens, T. (2024). How will the state think with ChatGPT? The challenges of generative artificial intelligence for public administrations. *AI & SOCIETY*, 1–12. https://doi.org/10.1007/s00146-023-01840-9

Casey, D., & Maley, M. (2024). Failing to Learn From Policy Failure: The Case of Robodebt in Australia. [Unpublished manuscript]. School of Politics and International Relations, Australian National University.

Chandra, Y., & Shang, L. (2019). *Qualitative research using R: A systematic approach*. Springer Nature Singapore. https://doi.org/10.1007/978-981-13-3170-1_8

Commonwealth. (2023). *Question 1923 - Shoebridge, Senator David to the minister representing the minister for defence*. https://parlwork.aph.gov.au/senate/questions/1923

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*, 228–239. https://doi.org/10.1080/14703297.2023.2190148

Dawson, P. (2020). Cognitive offloading and assessment. In M. Bearman, P. Dawson, R. Ajjawi, J. Tai, & D. Boud (Eds.), *Re-imagining university assessment in a digital world*, pp. 37–48. Cham: Springer.

Department of Industry, Science and Resources. (2023). *Safe and responsible AI in Australia—Discussion paper*. https://storage.googleapis.com/converlens-au-industry/industry/p/prj2452c8e24d7a400c72429/public_assets/Safe-and-responsible-AI-in-Australia-discussion-paper.pdf

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Elkhodr, M., Gide, E., Wu, R., & Darwish, O. (2023). ICT students' perceptions towards ChatGPT: An experimental reflective lab analysis. *STEM Education*, *3*(2), 70–88.

Evans, J. (2023, 26 May). Home Affairs experimenting with ChatGPT in refugee and cyber divisions. *ABC News*.

Falstein, M. (2023). *New South Wales Council for Civil Liberties Submission: Safe and responsible AI in Australia—Discussion paper*. https://consult.industry.gov.au/supporting-responsible-ai/submission/view/239

Firat, M. (2023). What ChatGPT means for universities: Perceptions of scholars and students. *Journal of Applied Learning and Teaching*, *6*(1), 1–22.

Foster, A. (2023). *Can GPT-4 fool TurnItIn? Testing the limits of AI detection with prompt engineering*. IPHS 300: Artificial Intelligence for the Humanities: Text, Image, and Sound. Kenyon College.

Fulbright, R., & Morrison, M. (2024). Does using ChatGPT result in human cognitive augmentation? *arXiv. arXiv:2401.11042*.

Li, G., Chang, L., Krebs, S., Zaidi, N., Whelan, C., & Doss, R. (2023). *Safe and responsible AI in Australia: discussion paper*. Submission by the Centre for Cyber Resilience and Trust (CREST), Deakin University. https://consult.industry.gov.au/supporting-responsible-ai/submission/view/414

Head, B. W. (2023). Reconsidering expertise for public policymaking: The challenges of contestability. *Australian Journal of Public Administration*. https://doi.org/10.1111/1467-8500.12613

HKU, T. U. o. H. K. (2024). *HKU ChatGPT*. https://its.hku.hk/services/university-wide-applications/hku-chatgpt/

Huang, J., & Huang, K. (2023). ChatGPT in Government. In K. Huang, Y. Wang, F. Zhu, X. Chen, & C. Xing (Eds.), *Beyond AI: ChatGPT, Web3, and the business landscape of tomorrow* (pp. 271–294). Springer.

Illingworth, S. (2023, 19 January). ChatGPT: students could use AI to cheat, but it's a chance to rethink assessment altogether. *The Conversation*. https://theconversation.com/chatgpt-students-could-use-ai-to-cheat-but-its-a-chance-to-rethink-assessment-altogether-198019

Lodge, J., Howard, S., & Bearman, M. (2023). *Assessment reform for the age of artificial intelligence*. TEQSA. https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

Marsden, C., Webb, G., Fawns, T., & McIntosh, P. (2023). *Safe and responsible AI in Australia*. https://consult.industry.gov.au/supporting-responsible-ai/submission/view/481

McDonald III, B. D., Hatcher, W., Bacot, H., Evans, M. D., McCandless, S. A., McDougle, L. M., Young, S. L., Elliott, I. C., Emas, R., & Lu, E. Y. (2024). The scholarship of teaching and learning in public administration: An agenda for future research. *Journal of Public Affairs Education*, *30*(1), 11–27.

Michels, S. (2023). Teaching (with) artificial intelligence: The next twenty years. *Journal of Political Science Education*, 1–12. https://doi.org/10.1080/15512169.2023.2266848

Migone, A., & Howlett, M. (2023). Assessing the 'forgotten fundamental' in policy advisory systems research: Policy shops and the role(s) of core policy professionals. *Australian Journal of Public Administration*. https://doi.org/10.1111/1467-8500.12595

Newman, J., & Mintrom, M. (2023). Mapping the discourse on evidence-based policy, artificial intelligence, and the ethical practice of policy analysis. *Journal of European Public Policy*, *30*(9), 1839–1859. https://doi.org/10.1080/13501763.2023.2193223

NIAA. (2023a). *First nations impact assessments framework*. https://www.niaa.gov.au/indigenous-affairs/closing-gap/implementation-measures/first-nations-impact-assessments-framework

NIAA. (2023b). *Priority reform four: Shared access to data and information at a regional level*. https://www.niaa.gov.au/2023-commonwealth-closing-gap-implementation-plan/changing-way-we-work/priority-reform-four-shared-access-data-and-information-regional-level

Ombudsman, N. (2023). *NSW Ombudsman submission—"Safe and responsible AI in Australia" discussion paper*. https://consult.industry.gov.au/supporting-responsible-ai/submission/view/357

St. Denny, E., & Zittoun, P. (2024). Introduction to the handbook of teaching public policy. In E. St. Denny & P. Zittoun (Eds.), *Handbook of teaching public policy* (pp. 1–15). Edward Elgar Publishing.

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, *6*(1), 1–10. https://doi.org/10.37074/jalt.2023.6.1.17

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, *76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

Turnitin. (2023). *Turnitin's AI writing detection capabilities—Frequently asked questions*. https://in.turnitin.com/products/features/ai-writing-detection

Weatherall, K., Bednarz, Z., Bello y Villarino, J.-M., Burgess, J., Cellard, L., Cohen, T., Fraser, H., Goldenfein, J., Graham, T., Haines, F., Henman, P., Ilyushina, N., Kennedy, J., Scully, J. L., Leeftink, D., Maitra, S., Matulionyte, R., McCosker, A., Mullins, R., … Zeng, J. (2023). *Safe and responsible AI in Australia discussion paper—ADM+S submission*. https://consult.industry.gov.au/supporting-responsible-ai/submission/view/437

Whiteford, P. (2021). Debt by design: The anatomy of a social policy fiasco—Or was it something worse? *Australian Journal of Public Administration*, *80*(2), 340–360. https://doi.org/10.1111/1467-8500.12479

How to cite this article: Casey, D. (2024). ChatGPT in public policy teaching and assessment: An examination of opportunities and challenges. *Australian Journal of Public Administration*, 1–15. https://doi.org/10.1111/1467-8500.12647

APPENDIX

ASSESSMENT TASKING DOCUMENT

Applied Policy Project (POLS3041) **Semester 2, 2023**

Due date and value: Please consult the assessment tab on Wattle

Word limit: 1000 words. Consistent with CASS word limit guidelines, there is a 10% leeway. Anything over that 10% will incur a flat 10 percentage point deduction. The word count will exclude any bibliography, attachments, or footnotes that do not provide any additional content.

File naming conventions and formats:

1) Your self-reflective essay. This is the only part that is subject to the word count. Please ensure that this document is named "SURNAME—TOPIC". This is the only part that the word limit applies to.
2) Your final brief, produced by ChatGPT/Bard. For this, you can cut/paste from ChatGPT/Bard, or provide a screenshot. Please ensure that the document is named "SURNAME—Final Brief—TOPIC."
3) Your full ChatGPT/Bard script. You can choose to submit a URL (in a Word/PDF doc) that ChatGPT generates or provide screenshots. Whichever works for you. Please ensure that the document is named "SURNAME—ChatGPT transcript—TOPIC."

All documents must be in either Doc or PDF formats.

Exercise 3: Summary Policy Brief and Self-Reflective Exercise

The third assessment builds on the first two. Your Deputy Secretary has agreed to your recommendation from exercise 2. You must now get ChatGPT (or similar) to write a 700-word policy brief, recommending that the Minister agree to your preferred approach. This includes getting

ChatGPT to include a First Australians Impact Assessment statement in the brief (First Nations Impact Assessments Framework | National Indigenous Australians Agency (niaa.gov.au)).

Students will then analyse ChatGPT's answer; the process of getting ChatGPT to write the answer; and a personal reflection on it. The personal reflection could consider either or both of the issues below:

1) from a policy studies perspective:
   - How did you get it to make the recommendation that you wanted? Did ChatGPT "agree" with you? Or did you need to "tell" it to make a certain recommendation?
   - As you interacted with ChatGPT, how much did its recommendations change? Why?
   - Is it politically responsive, and relevant to current Australian political circumstances?
   - How did you ensure the answer was factually correct?
   - What was it missing that you consider is important?
   - Is it a "useful" tool in policy development?
2) from an assessment perspective
   - Reflect on this task, as a piece of assessment. What have you learnt? Based on this experience, how useful do you think ChatGPT is for policy? For university assessment? How much work was it to get ChatGPT to produce a "good" piece of work? Or did you need to have to do so much pre/post work it wasn't worth it?

Students should:

- **Retell**: Describe what you did in the ChatGPT exercise
- **Relate**: Relate what you did to theoretical concepts and ideas in policy studies, political science, public administration, politics, and emerging ideas around AI and ChatGPT (this is deliberately broad, to allow you to bring in ideas from throughout your degree, not just this course). This requires you to reference existing academic and grey/popular literature (including on ChatGPT) to compare/contrast your experience.
- **Reflect**: Interpret your experience by reflecting on how the empirical material (your "retelling") links to the broader concepts and ideas.

Only the self-reflection essay counts towards the word count, but students must submit both the policy brief produced by ChatGPT and their entire ChatGPT transcript.

Refencing: Any footnote-based referencing style is acceptable.

Tips: I place a lot of emphasis on the quality of the writing. No matter how good your arguments are, readers will be distracted by spelling errors, grammatical errors, etc. Spend the time proof-reading your work. This includes get the name of organisations correct (e.g., Is the ATO the "Australian Tax Office" or the "Australian Taxation Office"?)

There is no suggested format. However, you may wish to consult this resource: Reflective Essays - ANU

In class ChatGPT exercises

In groups—ask these separately, compare and check results:

- Some maths—start simple then get complex (e.g., 342 × 78,932)
- A poem/song about something/someone
  - Maybe based on an existing song?

- Proofread some text
- Five scholars on public policy theory
- Five key journal articles on public policy theory
- Can you get it to take a political position on something?
- "The cow jumped over the"
- Can you get it to generate a "fact" that you know is accurate about your policy topic
- Different languages?

Details of the focus group

A few days after results for the course were released, I contacted the seven students that had agreed to participate in this research to organise the focus group. The focus group was held approximately 2 weeks later (19 December 2023) on Zoom. Five students participated in the focus group.

The indicative list of questions that guided the focus group discussion were as follows:

Question	Source/inspired by:
How did the assignment go?	
Describe your experiences using ChatGPT during the exercise? Did you find it enjoyable and helpful?	"Describe your experience using ChatGPT during the tutorial tasks. Did you find it enjoyable and helpful?" (Elkhodr et al., 2023, p. 74)
Was this your first time using ChatGPT?	Not explicitly asked, but discussed in Elkhodr et al. (2023)
What impact do you think ChatGPT will have on policy making?	New.
How did you assess the quality of ChatGPT's work?	Inspired by Tai et al. (2018)
What do you think you got out of/ learnt from the assessment?	Inspired by Firat (2023).
How did using ChatGPT compare to how you would have completed the exercise "normally"?	"How did using ChatGPT compare to using search engines for completing the tasks?" (Elkhodr et al., 2023, p. 74)
What skills do you think this exercise helped you develop?	Inspired by Tai et al. (2018)
What do you think ChatGPT means for university teaching and assessment? On your learning?	"What does ChatGPT mean for students and universities?" (Firat, 2023, p. 60) and "Impact on assessment and evaluation" as a theme coming out of Firat (2023).

The focus group was recorded and auto-transcribed. I then manually checked the transcript and provided it back to the participants to give them an opportunity to check it for accuracy. The transcript was manually coded to identify themes and issues and ChatGPT was also used to identify themes.