

Oxford SU Report on the Rollout of ChatGPT Edu at Oxford University

May 2026

By Jacob Hougie, Policy Associate at Oxford SU

EXECUTIVE SUMMARY	6
INTRODUCTION	7
CONTEXT AND LITERATURE REVIEW	8
AI and LLMs	8
GenAI and LLMs at Oxford	10
Generative AI and Teaching	11
RESULTS AND KEY INSIGHTS	13
Methodology	13
Weaknesses	14
Teaching, Learning and Inequalities	14
How often do students use ChatGPT Edu	14
How do students use ChatGPT Edu?	15
What are the implications of ChatGPT Edu use among students?	17
What do students think of teaching staff using ChatGPT Edu?	18
University Approach	19
Policy Clarity	19
Integration of GenAI into Teaching	20
Summative Assessment	21
Privacy and ChatGPT Edu	22
Environmental and Social Consequences of GenAI at University	22
Technical Details	23
CONCLUSION	23
BIBLIOGRAPHY	26
AI DECLARATION	28

Executive Summary

1. This academic year (2025/26), Oxford University began offering ChatGPT Edu licences to all students. This was intended to provide students equal access to advanced Large Language Models (LLMs) to support their learning, as well as improving privacy and data security. Licences were also granted to staff to improve teaching and administrative efficiency.
2. This report uses a survey (n=35) and focus groups to improve understanding of the consequences of this rollout and provide input into University's evolving approach to Generative AI (GenAI). Our survey received a smaller response than hoped, but by using its findings alongside other surveys and the detailed qualitative research that accompanied it, this report still provides significant impact.
3. Broadly, we found that students have welcomed ChatGPT Edu, but this is limited by perceptions of policy as weak and there is a risk of inequalities developing over the long-term.
 - a. Oxford students are using GenAI carefully, which could have significant educational and equity benefits. But our research suggests that some of the ways they are using it risk creating inequalities, especially for students with SSPs, for those whose first language isn't English, and around AI skills. Further investigating these findings could avoid significant future problems with the rollout of ChatGPT Edu.
 - b. Students think University policy is unclear and inconsistently applied, while GenAI is poorly integrated into teaching. This limits the impact of ChatGPT Edu, creates a degree of fear around GenAI use and appears to punish students arbitrarily. Integrating GenAI into teaching and assessment could direct students to more effective GenAI use, improve clarity around GenAI policy and communicate that policy more effectively.
4. Key findings include:
 - a. Oxford students seem less likely than students nationally to use AI generated content without editing it, and they also seem to engage cognitively with their learning rather than relying on GenAI.
 - b. AI could reduce the impact of the 'hidden curriculum'. It may also help students whose first language is not English, but the limitations of LLMs may reduce this benefit.
 - c. It seems that students with SSPs are less likely to use ChatGPT Edu, limiting its benefit to them. There may also be inequalities between students' AI skills, depending on who chooses to develop them.
 - d. Students are cautious about teaching staff using GenAI, but especially given concerns about pay, support some use of GenAI by staff. The key is that the unique subject expertise of teaching staff is retained, so that university teaching retains an edge over what an LLM provides.
 - e. Many students think purchasing ChatGPT Edu suggested a less hostile response to GenAI use in education, but most still think policy and institutional attitudes are too vague and inconsistently applied for them to feel confident that they will not be accused of GenAI use. They particularly object to AI disclosure forms and a focus entirely on in-person exams, which treat everyone as guilty.

- f. 17% of respondents to our survey thought that ChatGPT Edu is less private than other LLM interfaces, despite the University's public claim that its agreement with OpenAI protects data privacy.
 - g. A broad range of students seem to share concerns about social and environmental consequences of GenAI, but they broadly see these as matters for regulation.
- 5. While there is much good news in this report, the report makes the following recommendations to mitigate against more negative outcomes:
 - 1. Further investigate whether and why students with SSPs are not using ChatGPT Edu as frequently as other students.
 - 2. Monitor inequalities between students, especially between those developing and not developing AI skills, as well as between those with and without English as a first language.
 - 3. Clarify departmental policy and integrate policy into teaching so that it is effectively communicated to students and staff.
 - 4. Review the use of AI disclosure forms.
 - 5. Integrate GenAI into teaching, perhaps using custom GenAI tools, to direct students towards ethical and effective use of GenAI, as well as help students to develop judgement with GenAI tools.
 - 6. Use a diversity of assessment methods, considering how different assessment methods promote learning, as well as their collective reliability.
 - 7. If possible, publish data privacy agreements with OpenAI to increase trust in ChatGPT Edu.
 - 8. If possible, increase usage limits, allow transfer of conversation history from personal ChatGPT accounts to ChatGPT Edu accounts, and secure agreements for access to a diversity of LLM interfaces.

Introduction

- 6. Artificial Intelligence (AI), especially Generative AI (GenAI) and Large Language Models (LLMs) have become ubiquitous in the last 3 years. The idea of AI has spread across our collective consciousness, its outputs across social media, and interfaces to interact with LLMs are open on more and more screens as one walks around. Indeed, it has the appearance of a technology as impactful on the economy and our lives as the personal computer, smartphone or the internet.
- 7. It is also a technology that appears particularly consequential for Higher Education institutions like the University of Oxford. It poses significant challenges to universities' key functions, including their educational and research functions, their accrediting function and their role in preparing young people for employment. GenAI offers a substitute for universities' educative function. It is capable of generating text to impart knowledge, explain facts and theories or provide analysis, mirroring elements of the teaching universities offer and the research they generate. Universities have a role in providing accreditation, professional or otherwise, to students that have demonstrated competency in activities relevant to their degrees or other courses. These accreditations are used to determine if someone should be supported in further study or employment. Universities have always had to ensure the reliability and accuracy of their accreditation and to prevent, identify and punish cheating. However, LLMs offer a new way of cheating which seems

both much easier and harder to detect than other methods. Finally, universities don't just have a role in educating students in their subject matter and discipline-specific skills, indeed they also have a role in developing transferable skills that will be useful for employers. LLMs may change the nature of work, but what work will involve in the short-to-medium term is not yet clear. This makes it harder for universities to understand which transferable skills to prioritise. Yet, GenAI also creates new opportunities for universities – to speed up administrative processes, make staff more efficient and improve teaching. Taking advantage of GenAI while minimising risks will be essential for universities in the near future.

8. In responding to GenAI, universities need to understand their students because the effects of any given policy will depend on their students. Understanding why and how students use GenAI can help them to understand the impact of existing policies and the way students might behave under new policies. Nationally, surveys suggest that almost all students are using GenAI to do something that contributes towards assessed work. It is also controversial, with 50% of students in a recent survey reporting that its existence has improved their experience of university and 16% reporting that it has made their experience worse. Students highlight GenAI's ability to improve their efficiency and their learning, but others highlight its societal and environmental impact, as well as raising concerns about assessment and their ability to develop academic skills using GenAI (Stephenson and Armstrong 2026).
9. Student perspectives are also tightly associated with a number of concerns arising from the qualities of GenAI itself: its environmental impact, its impact on global and local inequality, the risk of false and misleading outputs, the ease with which GenAI can be used to generate misinformation and the potential for people's cognitive ability to degrade when using GenAI.
10. The University of Oxford has already begun to respond to GenAI. Most significantly, it has purchased licences for all staff and students to access ChatGPT Edu, a special interface for OpenAI's ChatGPT chatbot based on their GPT models. This report covers student perspectives on GenAI but has a strong emphasis on the rollout of ChatGPT Edu at Oxford.
11. The report will begin with a literature review, explaining how LLMs, like the model behind ChatGPT Edu, function and what their properties are in general, then exploring the existing literature on GenAI in Higher Education. The literature review grounds the key questions the report explores when it turns to its primary research findings. Through a survey sent to all students, which received 35 responses, and focus groups involving 10 students selected after an expression of interest sent to all students, the report builds an understanding of how students use ChatGPT Edu, its implications of their learning and their wider attitude to LLMs. In discussing these insights, the report will draw out the key points of feedback to support the University in developing its approach to GenAI.

Context and Literature Review

AI and LLMs

12. LLMs are a complex form of machine learning. They differ from other methods primarily in their scale – they rely on mammoth data sets to train models that are very large in terms of the set of mathematical weights that result from training.¹

¹ For early concerns about LLMs and the implications of scale, see Bender et al. (2021).

13. During training, they are provided with large quantities of text. Text is broken down into fragments of words, called 'tokens.' The model is trained to predict the next token in a sequence, given the tokens that come before it. In doing so, it learns how tokens relate to one another: which tokens appear near each other and how they inter-relate. With enough data, an LLM can process new text and break down, for example, which noun a pronoun refers to, and then use the start of the text to predict which tokens are most likely to come next. Then, when an LLM is trained and it is provided with new text it can understand the most probable next tokens to follow up. Crucially, they do not just provide each token in turn, but look forward to produce a set of tokens which are most probable in combination (Murgia 2023).
14. LLMs are technically impressive for their ability to produce human sounding text, but they have significant weaknesses as a form of machine intelligence.
 - a. Hallucination – at some level, LLMs can be understood as prediction machines, so they just put together words in a way that sounds correct. When an LLM has been trained on enough data to give it true things to say, it can produce text that aligns with the truth, since the truth is more like the text it has seen before. But when it hasn't, it can produce text that sounds like the sort of thing that might be true without being true (Kalai et al. 2025). This is why it can produce academic references, correctly formatted, with article titles, journals and authors that sound right and could go together, but which don't exist.
 - b. Bias and homogenisation – since LLMs will generally pick the most probable batch of tokens, it will often pick the same or similar groups of tokens. This might be troublesome simply because it homogenises outputs, but it is especially troubling because of its sample data. We don't know exactly what LLMs are trained on, but they seem to be trained mainly on Western sources, and it is very hard to remove the biases internal to that culture and the text it generates. Worryingly, LLMs will largely take the most likely option, not the second most likely. For example, if the training data includes bananas much more than oranges, it would almost always say banana when asked for a fruit, rather than saying banana in proportion to how much it occurs in the dataset. This leads to very similar outputs, all in a Western framework, which all include similar biases (Ananya 2024; Nguyen et al. 2023; Dennison 2025; Qian 2025)
 - c. Energy and water – training and using an LLM is a computationally intense process, requiring very significant amounts of energy to power and cool the necessary computers, as well as requiring lots of water, drawing on local water supplies, to cool them. Many data centres around the world are also dependent on polluting energy sources, contributing to climate change and worsening local air quality (Reitmeier and Lutz 2025).
 - d. Sycophancy and mental illness – LLM-based chatbots have a strong tendency to agree with the user. Evidence suggests users prefer this, which reduces the incentive for LLM-providers, who are ultimately selling a product, to reduce this tendency. However, it generates a significant risk of encouraging wrongful, harmful or delusional thoughts and behaviour. This poses a particular risk to those living with mental illnesses, whose anxieties may be enabled and exaggerated through LLM use (Cheng et al. 2026; Hall 2025).
 - e. Data collection – LLM-providers need more data to improve their models. There is therefore also a risk that LLM-users will provide personal data, in the form of their

prompts to and conversations with an LLM, which could be handled irresponsibly leading to data breaches.

- f. Jagged frontier – LLMs are not equally good at all things. Crucially, it is often hard to tell what an LLM is good at because it presents its responses confidently, as with a hallucination. This creates a risk of people over- or underestimating LLM's abilities (Lukes, n.d.).

GenAI and LLMs at Oxford

15. The University of Oxford is subject to these trends in much the same way as other universities, but there are some particular features of Oxford that are relevant context to the rest of this report.

- a. Its teaching system is particularly personalised through the small group teaching of the tutorial system. In education, part of what LLMs offer is the potential to replicate highly personalised teaching and feedback. The University's educational role is therefore especially threatened, and the University needs to show the distinct advantage of in-person, human tuition over LLMs.
- b. The University already uses more in-person exams than other universities. While the University had partially shifted away from in-person exams, the risk of LLM cheating has slowed and partially reversed that change. However, any further shift towards in-person exams will be constrained by capacity limits in the Exam Schools and because of the disadvantages of in-person exams as a form of assessment.
- c. The University is one of a number that have purchased licences for all students and staff to access to the most recent AI models. It purchased ChatGPT Edu licences for students in summer 2025. This will be the focus of the report. The intention behind this is to ensure equitable access to models and the learning support they provide, improve privacy and data security and to improve university administration and teaching delivered by staff. ChatGPT Edu comes with limitations to OpenAI's data retention, as well as unlimited use of some of the most advanced models, and limited usage of the absolutely most advanced models.

16. Alongside providing access to models, the university has established the AI Competency Centre to offer training to staff and students on how best to use advanced models. This includes online guidance as well as in-person and online seminars on using GenAI effectively. The University has also supported a variety of experiments in using GenAI to improve teaching, including using GenAI to provide course materials in a more accessible way for students with learning needs (AI Competency Cent., n.d.-b), and creating custom GPTs with additional instructions and knowledge bases to improve language learning (AI Competency Cent., n.d.-a).² The University is reviewing and researching aspects of both teaching and assessment to understand how best to respond to and adopt GenAI.

17. Right now, most policy is decided at a departmental level because GenAI poses different opportunities and challenges to each department, depending on the subject-matter and skills they teach. However, the University has centrally set that:

- a. In summative assessment, academics should consider how AI can be used, ensure equitable access to relevant tools, require AI declarations and only use approved AI detectors. Meanwhile, students should only use AI as permitted, and acknowledge

² The AICC hosts a database of these projects: <https://oerc.ox.ac.uk/ai-centre/using-ai-at-oxford/case-studies>.

their use in AI declarations. (University of Oxford 2025). At least within the Department of Politics and International Relations, AI declarations are sent to examiners, but only so that they can bear the content of the declaration in mind if they think summative work uses AI in a way that is not permitted.

- b. More generally, AI should only be used to support human work, that one should be transparent about AI use, and that one should be conscious of data security, hallucination, bias and legality.

Generative AI and Teaching

18. A recent survey of 1,054 students organised by the Higher Education Policy Institute suggests that 94% of students are using GenAI in at least one way that contributes towards assessed work. Mainly this involves explaining concepts, summarising readings or structuring their thoughts. 25% of students in the survey reported using GenAI to generate text for their assessed work which they then edited and 12% reported including text that was generated by GenAI and not edited at all by them. Students are to some extent relying primarily or entirely on GenAI rather than traditional sources. They do this mainly to save time, which they generally reinvest in studying, improving their work and learning and gaining skills in working with AI (Stephenson and Armstrong 2026).
19. Universities in the UK have adopted a variety of stances to GenAI. The Russell Group of universities adopted principles in 2023 that included encouraging appropriate use, adapting teaching and assessment to AI use and encouraging academic rigour. These principles are broadly reflected in member university's policies (Wilson 2025). Survey evidence among students also suggests that more universities, especially Russell Group universities, are encouraging GenAI use (Stephenson and Armstrong 2026). More universities are also purchasing licences for GenAI tools for their students.
20. A small number of universities have also developed custom GenAI interfaces for staff and students.
 - a. Edinburgh's ELM system is a custom interface to access OpenAI's GPT models (with a data security agreement like Oxford's) but also includes access to a version of Meta's Llama 3.3 model hosted on the university's servers (which it claims are more energy efficient). This interface allows the university more influence over how students use LLMs (Univ. Edinb. Inf. Serv. 2024).
 - b. Arizona State University developed 'CreateAI', which allows staff to create course specific AI tools designed to promote learning on specific courses. This includes tools to help staff develop custom quizzing, scaffolded learning and writing feedback tools, among others (Ariz. State Univ., n.d.).
 - c. Other universities have also, like Oxford, trialled custom GPTs as a tool for improved GenAI use in learning (Rasovic and Svingen 2025).
21. Students *are* using GenAI, but the implications of that fact are much less clear. The education impact is one of the more significant questions and research offers a mixed picture and it seems very heavily on how GenAI is used and in what context. For this report, one meta-analysis was consulted, a systematic review and several individual studies were considered in detail.
 - a. The meta-analysis found that most studies reported a positive impact on academic performance following ChatGPT use but caveated this with the fact that many of the studies allowed, or did not state whether they allowed, the use of ChatGPT when academic performance was assessed as well as during teaching. It also

found that some subjects saw a positive impact, whereas others had no impact, and that studies with a duration of less than one week had no impact. The positive impact was explained by more personalised learning, immediate access to information and diverse framing, or allowing deeper engagement. They also found that most studies showed reduced cognitive load while using ChatGPT and that when there was reduced cognitive load, ChatGPT had a negative effect on academic performance and vice versa (Deng et al. 2025).

- b. The systematic review covered the key potential uses of GenAI – automated, detailed and personalised feedback to improve work, support with learning, brainstorming, practice questions, and to develop key skills. However, it raised the risk of hallucinations teaching things which are not true, the reproduction of bias, limitations of students lack AI skills and risks from students becoming too dependent on GenAI (Qian 2025).
- c. A randomised controlled trial at Harvard tested an AI tutor for physics against human tuition, controlling for pre-tuition ability. They created a custom AI tutor that was given a system-wide prompt to encourage active learning and manage cognitive load, was placed in a user interface that scaffolded learning, and was given correct answers to ensure accuracy. They found that students who received AI tuition from this custom AI performed substantially better when tested without access to AI (Kestin et al. 2025).
- d. Another randomised trial in Ghana used a non-modified version of ChatGPT and tested two groups of students on creative, critical and reflective skills before and after three weeks of teaching, with or without ChatGPT. Both groups watched lectures and prepared for discussion questions during the three weeks, but the test group used ChatGPT to prepare for the discussion questions. At the end, the group that used ChatGPT had significantly improved their scores, and in interviews described improved grasp of material (Essel et al. 2024).
- e. On the other hand, another randomised controlled trial asked two groups to conduct research on a public health topic and provide recommendations. One group used traditional search engines, the other used LLMs. They found that students that used LLMs experienced lower cognitive load and provided lower quality justifications for their recommendations (Stadler et al. 2024).
- f. Another randomised controlled trial, which has not yet been peer reviewed, split high school students in Turkey into three groups. All groups were taught a maths topic by a teacher and were then asked to revise and do practice questions. One used traditional study materials, one used an interface very similar to GPT-4, one used a custom version of GPT-4 similar to the Harvard study that was designed to provide better tuition. The students who used the custom version of GPT-4 did much better in practice questions, but no better than the control in the exam, and those who used the standard version of GPT-4 did somewhat better in practice questions but performed 17% worse in the exam. Notably, the students using the custom version of GPT-4 were much more confident in their performance. The key issue was that students using the standard version of GPT-4 did not benefit from the practice questions as much because they mainly asked for the answers, rather than for help, and did not engage with the material (Bastani et al. 2024).
- g. This suggests that some of the key moderating factors between academic performance and LLM use are cognitive engagement with material, which is supported by custom versions of LLMs designed to encourage such engagement.

It may also be relevant to ensure that LLMs do not provide incorrect advice or hallucinate answers by providing model solutions. That said, these studies are only preliminary, may be misidentifying causality, may be dependent on their context, how long LLMs are used for and the academic ability of the students.

22. There is also some evidence that GenAI use can help students to build judgement in using GenAI well (Qian 2025) and to help with administration and time management, with some evidence suggesting it may be particularly helpful for disadvantaged students (Edsights 2024).
23. Then there is the question of how to assess students given that, at various points in their learning, preparation for assessment or of assessed material, they may have access to GenAI. The Student Union's report of 2025 on undergraduate assessment covers some of these issues and provides student input.³ At some level there are three options: embrace AI and design assessments that permit AI use or even focus on how you use AI, use in-person exams to ensure that (even if preparation work involved AI) the tested output is definitely produced by the student, or use a combination of different assessments to balance things out, but also to ensure the integrity of single pieces of assessment. The key is to find a method of responding to AI that tests students for their competency in their degree and which can be reliably attached to a given student.
24. Another factor in choices about assessment policy is the potential student response. Students are particularly concerned about being accused of misconduct or assessments being made harder on the assumption that GenAI will be used (Stephenson and Armstrong 2026). In the former case, unclear policy can mean that students who are more cautious are disadvantaged relative to other students. In the latter case, such students would be especially disadvantaged relative to a pre-AI world. While making assessments harder can be a reasonable response, it cannot just punish those who are more fearful of punishment, or just create stronger incentives to cheat with AI.

Results and Key Insights

Methodology

25. To provide insight, this report used two primary research methods. First, an anonymous survey was sent to all students via the Student Union's all student email. 35 students responded, 57% of whom were undergraduates, 17% were postgraduate taught students and 16% were postgraduate research students, roughly corresponding to the proportions across the university though slightly overrepresenting undergraduates. 26% of respondents reported having an SSP and 69% reported not having an SSP, with the rest preferring not to say, though this does roughly correspond to the university-wide statistics. This survey included a mix of questions about how people use ChatGPT Edu and its integration with teaching as well one testing for how long and thought-out students' prompts are when using ChatGPT Edu. The latter presented students with three prompts at three levels of detail (one set for AHSS and Social Sciences students, another for MPLS and Medical Sciences students) and asked which was most like the prompts they use.
26. Second, 10 students were invited to focus groups. 25 students responded to an expression of interest form sent to all students alongside the survey. Invites were sent to try and ensure rough gender balance, a balance of levels (though PGR students were underrepresented among those who expressed interest) and of divisions (though MPLS was underrepresented among those who expressed interest). Two focus groups focused

³ See Hougie (2025) which was reported to the Taught Degrees Awarding Panel in June 2025.

on AI in teaching and learning, and one of those included someone who tutors, and the final focus group focused specifically on AI and ethics.

27. The rest of this report presents the findings of these two sources together.

Weaknesses

28. Our survey did not receive as many responses as we would have liked, just 35 responses out of over 27,000 students. It is very possible that this sample would include the extremes of GenAI use, with those most engaged with question about GenAI being the most likely to respond.

- a. The small size of our sample means that some of our claims may need further investigation to confirm. This problem becomes more significant when findings are broken down by division, or other factor, so these findings need to be taken with caution.
- b. However, it did receive responses from students who reported never using ChatGPT Edu, frequently using it and using it semi-regularly. This suggests there was some balance, but it may not be a representative balance.
- c. The survey responses also overrepresented MPLS students and underrepresented Social Science students. This may mean that certain uses of GenAI will appear more common in our sample than they are.
- d. We recommend reading this report alongside analysis of other GenAI surveys run by the university. If there are comparable questions between surveys, and our responses seem similar, that might confirm that our survey is representative to a similar extent. However, if those surveys also received low response rates or overrepresent certain groups, caution must still be taken.

29. Our focus groups were qualitative, but the underrepresentation of MPLS in our focus groups may help to balance out the overrepresentation of MPLS in the survey. That said, participants gave their own experience and cannot be assumed to be representative of the whole university.

Teaching, Learning and Inequalities

How often do students use ChatGPT Edu

30. The first thing to know about the rollout of ChatGPT Edu at Oxford is that most students seem to be making use of GenAI in some way. 69% of respondents reported to our survey report using ChatGPT Edu at least once a month and 77% of respondents used either ChatGPT Edu, or another LLM, at least once a month. Our sample size was small so these figures may undercount or overcount ChatGPT Edu users, but this suggests that a significant majority are using LLMs. 45.7% of respondents reported using ChatGPT Edu more than once a week or daily, suggesting that many students use ChatGPT Edu quite regularly. These findings should be considered alongside the findings of other university research.

31. Concerningly, in our survey, ChatGPT Edu usage appeared to be split by whether students reported an SSP, though only 26% of respondents reported an SSP so this is a very small sample. Among student with an SSP, 55% reported never using ChatGPT Edu, compared to just 21% among those who did not report an SSP (Figure 1). One putative benefit of LLMs is their ability to provide educative content in ways that support students with particular learning needs (Harkins-Brown et al. 2025). If students with these needs aren't using ChatGPT Edu, students are less likely to benefit, but it may also be worth investigating why such students aren't using ChatGPT Edu.

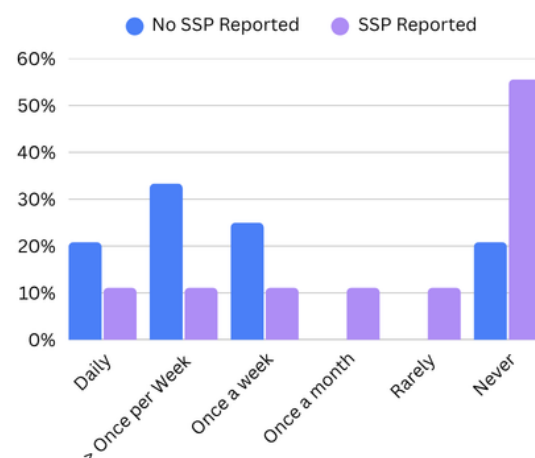


Figure 1: Frequency of ChatGPT Edu use by reported SSP. 9 students reported an SSP, 24 did not, 2 preferred not to say (not shown).

32. In our survey, it is also notable that 20% of respondents used other LLMs more frequently than ChatGPT Edu and 14% used other LLMs significantly more frequently than ChatGPT Edu. Some reasons for this came up during focus groups: that Claude is more useful for programming, that they have personal ChatGPT subscriptions that include their chat history or higher usage limits, or that they use other interfaces, particularly NotebookLM because one can upload files to increase reliability.

33. Please note that all further results refer only to students who reported using ChatGPT Edu at least 'rarely.'

Recommendation 1: Further investigate whether and why students with SSPs are not using ChatGPT Edu as frequently as other students.

How do students use ChatGPT Edu?

34. Students at Oxford use ChatGPT Edu for a variety of tasks. The most common tasks reported in our survey were: 'advice on specific topics or readings' (82% of respondents), 'general advice on learning' (63.27%), reviewing written work (51%), and summarising readings (40%). Less common tasks included: careers advice (31%), life advice (21%), generating formative work (21%), generating summative work (20%). One weakness of this report is that coding was not included as a task separate to generating formative/summative work so this is not shown distinctly (Figure 2).

35. It is hard to directly compare these findings to national findings in the Higher Education Policy Institute's 2026 survey, since this survey used different phrasing. However, they found that 25% of students used GenAI to generate text for assessed work, with additional editing, 18% did so with editing by a digital tool and 12% without any editing (Stephenson and Armstrong 2026). In our survey, 1 of the 5 students who reported using ChatGPT Edu for their summative work at Oxford explained in a comment that this was a task they were set and 2 of 5 offer answers at other points in the survey that suggest they rarely wholly accept answers generated by ChatGPT Edu. This suggests that only 2 of 35 students who responded to the survey (5.7%) submit work generated by ChatGPT Edu without editing

it. This is significantly lower than the 12% in the national survey, though our survey may not be representative.

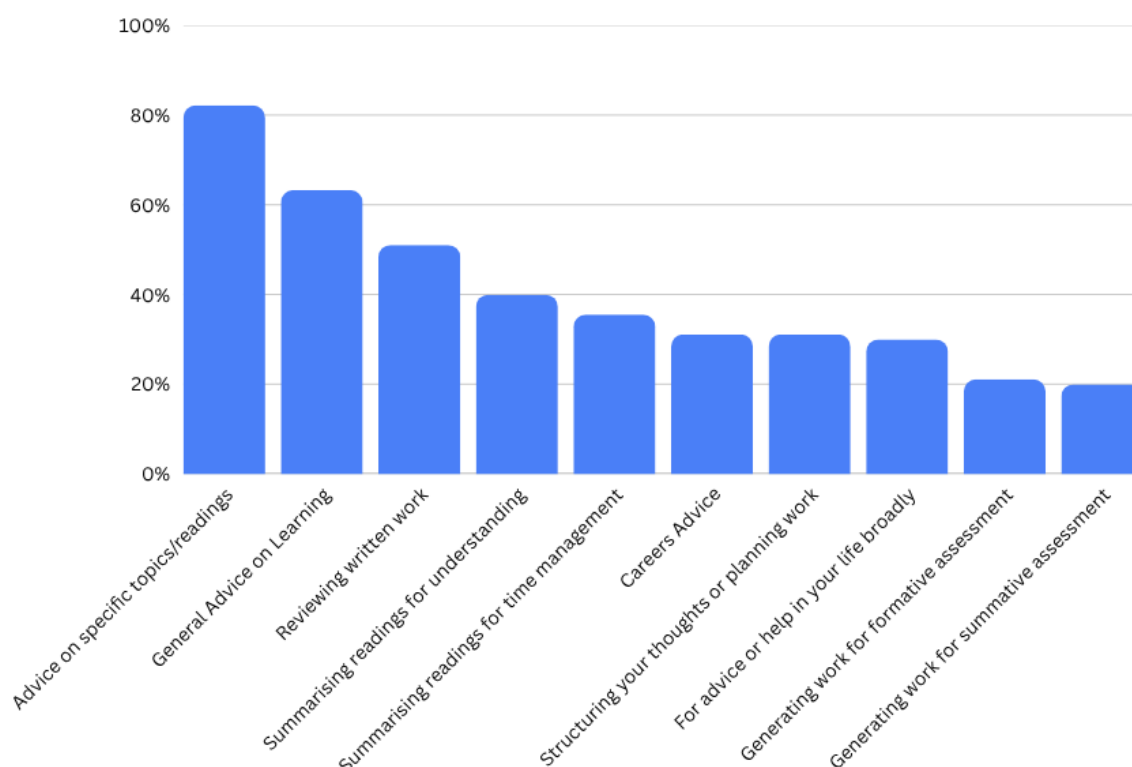


Figure 2: Frequency of specific uses of ChatGPT Edu as a percentage of total ChatGPT Edu use. 30 respondents answered this question.

36. Our data also suggests that students think carefully about how they make requests to ChatGPT Edu and how they use its outputs.

- a. In one question students were presented with three different prompts. The first included (a) context, (b) a list of specific requests and (c) a summary of an understanding of the answers to those requests. The second included (a) and (b), but not (c). The last included only a general request. Only 17% of students asked for the final, least specific prompt. This suggests that students are thinking carefully about what information they want from ChatGPT Edu, are thinking about topics before going to an LLM for specific requests and/or how to get better outputs by providing context and specific requests.
- b. In another question, students were asked how often, and to what extent, they reject the outputs of ChatGPT Edu. Most often, students partially reject ChatGPT Edu's outputs, though they rarely wholly accept its outputs. Generally, students reported varying how much they accepted ChatGPT Edu's outputs which suggests that they thought and considered the results before acting on them (Figure 3).

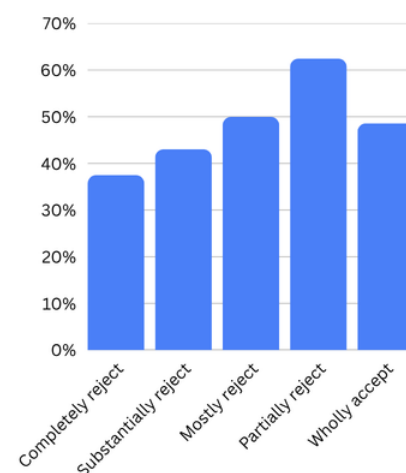


Figure 3: Frequency of extent of rejection of ChatGPT Edu outputs, as a percent of ChatGPT Edu use frequency. 30 respondents.

- c. Voiena et al. (2026) identify three risks of LLM use in education: (a) outputs appears complete and appears not to need interpretation (b) LLMs can do all steps of working and so don't require agency which prevents learning since you do not need to engage with material and (c) humans use fluency as a proxy for correctness which can lead us to assume LLM results are correct. If students think carefully about what they put into LLMs, what it outputs and how it gets there, these three risks can at least be diminished.

What are the implications of ChatGPT Edu use among students?

- 37. Students are using LLMs to make themselves more efficient. In our focus groups, most students described using ChatGPT to prepare revision materials more quickly, to work out what readings were most worthwhile (which they compared to reading an abstract, introduction or conclusion), or checking their work, especially programming work.
- 38. However, many of the same students described feeling they had to use LLMs in order to keep up with their course loads and to avoid falling behind other students who were using LLMs. One student described a collective anxiety among non-ChatGPT users that, after the University bought ChatGPT Edu licences, 'we'd sort of be cheated disadvantaged by the ones of us that were using ChatGPT', expanding that they need ChatGPT Edu to keep on work. Another expressed a concern that 'if you don't use ChatGPT you will just fall behind because everyone else will be like flying and soaring through things.'
 - a. On the one hand, this might be concerning – if students feel like they need to rush through their learning, they may use LLMs in ways that are detrimental to their learning.
- 39. On the other hand, if they can use LLMs to spend more time doing their work, they can engage more deeply with materials and genuinely. This might mitigate preexisting concerns about workload at Oxford, giving students new tools to manage workload.⁴ Those who struggled most with their work could manage their work better and fall below any critical threshold at which work becomes unmanageable.
 - a. If this benefit is realised without harming learning outcomes, this could be a major benefit to learning equality at Oxford. This report is not enough to draw conclusions, however, so the impacts should continue to be observed.
- 40. It seems unlikely that students will develop AI skills given the current state of students' intentions with ChatGPT Edu, raising the risk of potential inequalities. In our survey, 66% of students reported that they had not looked at any resources on the effective use of AI, 24% said they had only read the AICC's web resources and 10% said they had both read the AICC's web resources and attended workshops or webinars. This suggests that just one third of students are engaging with resources on the effective use of AI, with an even smaller group engaging intensely with these resources.
 - a. In our focus group we also found that only some students use ChatGPT Edu because they want to get better at using LLMs, while other students did not value AI skills, with one student saying they 'admire [those who avoided LLMs] more that they're able to get their work done', rather than respecting those with strong AI skills.
 - b. On the other hand, students in the focus group reported learning from friends and through experimentation to get better with AI. All the same, this represents a

⁴ See for example (Chapman 2024).

potential risk of inequality in how effectively students use AI and how much they develop their skills in AI use (if these are indeed valuable skills).

- c. One noteworthy finding of our survey was that postgraduate students reported being slightly more confident in their work after using ChatGPT Edu than undergraduates. Undergraduates reported, on average, being exactly as confident in their work after using ChatGPT Edu (3 out of 5 on a scale where 5 meant more confident and 1 meant less confident). By contrast, postgraduate students reported being on average 0.5 points towards more confident after using ChatGPT Edu. This may suggest that students who are more advanced in their academic careers can feel more confident after using ChatGPT Edu.
 - d. Social Science and Medical Science students also reported higher confidence (3.5 and 4.25 respectively), but these are sample sizes of 6 and 4 respectively so we can have low confidence in this.
41. Students are also conscious of other potential equity implications of LLMs. One considerable upside of LLMs is their potential to limit the effects of the 'hidden curriculum.' This concept refers to the idea that some cultural capital or study skills knowledge is more available to certain groups of students than other, and this this hinders academic progress (Universities UK and National Union of Students 2019; Mountford-Zimdars et al. 2015; de Henau 2024; Li 2024). In our focus group, students mentioned that LLMs help them to understand concepts not included in lectures, or to explain parts of lectures they did not understand, to guide their approach to their learning, and to adjust to the particular academic standards of UK higher education. This also fits with our survey, which suggested that a large proportion of students are using ChatGPT Edu to guide their learning more broadly. Once more, the impacts of these changes need to be monitored more closely, but there is a significant potential upside from this.
42. There are however two potential limitations of this. First, GenAI might be biased towards certain epistemologies and worldviews (Tao et al. 2024). While this may be a concern for students who wish to use this perspective, students mainly framed this as a social issue, rather than pedagogical, and were concerned for potential global inequalities. Second, LLMs perform worse in other languages because they lack the same volume of training data that makes them effective in English (Pava et al. 2025). In our focus groups, students simultaneously noted the usefulness of ChatGPT Edu for students whose first language is not English to translate and improve their English, while caveating this with the risk that it would be less useful if they wished to use their own language. Exactly how this balance will play out will also need to be watched closely.

Recommendation 2: Monitor inequalities between students, especially between those developing and not developing AI skills, as well as between those with and without English as a first language.

What do students think of teaching staff using ChatGPT Edu?

43. Finally, there is the question of what students see as acceptable use of LLMs by staff. This was not covered in our survey, but was a major focus of one focus group, which included a DPhil student who themselves also ran tutorials.
44. Students agreed that tutors or examiners using LLMs to mark work would be deeply uncomfortable, since it is important that human judgement is involved in the process, but

beyond that it got more ambiguous. Students in the focus group agreed that it was acceptable, or good, for tutors to use LLMs to structure and plan tutorials more efficiently and effectively. This was especially because students recognised that tutors are only paid for time in tutorials, not for preparation. On the other hand, students did not think it was acceptable for staff to prepare lecture materials using LLMs since it was considered important that they themselves, as a subject experts, bring some extra to their teaching that students couldn't just get themselves from LLMs. That said, there was still some debate about the extent to which this was acceptable and whether some use of LLMs for general topic introductions might be acceptable.

University Approach

45. Throughout our focus groups a consistent concern arose about the university's approach to GenAI, which students found unsystematic, confusing and at times arbitrary. According to our survey, 80% of students had read either the University, or their department or their colleges' advice on the ethical use of AI, but in our focus groups, most students reported only a vague level of familiarity with university AI policy. Overall, our research suggests a significant need for the university to develop a more coherent strategy for GenAI in teaching and assessment that encourages students to develop judgement about how to use GenAI

Policy Clarity

46. Students welcomed the purchase of ChatGPT Edu licences and felt like it communicated to students that the university was taking a pragmatic approach. However, they also described the policy as 'volatile' after academics had previously emphasised the importance of not using GenAI. They also felt that, even after the purchase of ChatGPT Edu licences, policy had not been clear, creating a sense that even while GenAI was officially permitted, one could never be sure if you were using it in a way that was institutionally accepted, or accepted by the particular members of staff you interacted with. This confusion was mixed with a sense that since students now all had access to ChatGPT Edu, one needed to use GenAI to avoid falling behind other students. This further muddled students' sense of what the university expected of them and what they thought was best for their learning.
47. A particular concern was the requirement for AI disclosure forms. Students felt that no student would ever knowingly admit to inappropriate use of GenAI in these forms, but that even if you were confident you had followed policy these forms felt like 'an admission of guilt'. These further contributed to a sense that the university disapproved of GenAI use.
48. Students were worried that they'd misunderstand policy, accidentally admit to it and be punished. One said they would be 'scared to say [what I'd used GenAI for] even though I don't think I've done anything wrong.' Another student said that even though they knew all their peers used GenAI in some way, no one would ever admit it to a tutor.
49. This fear was exacerbated by stories, shared among friends, of arbitrary applications of policy for formative work. In at least one department, policy states that staff do not need to mark work that they suspect to be produced using GenAI in an unacceptable way. One student in that department shared a story of a friend who did not use GenAI being told in a tutorial that their work would not be marked because of suspected GenAI use. Without any right to appeal, this policy allows for staff to arbitrarily judge that work does not need to be marked, contributing to a sense that accusations of GenAI use will be dealt with in an arbitrary manner, contributing to fear about AI policy.

50. Ultimately, students think the university policy is not clear. This may be because of a low level of familiarity with the policy, but either way work needs to be done to ensure students feel confident about what they are allowed use GenAI for and to feel confident that they will not be punished for doing so.
51. Of course, there remains the question of what content should be in policy once it is made clear. On this, students recognise that educationally appropriate GenAI use will depend on the department, level and course and what skills a student needs to develop. As such, policy needs to be clearer while at the same time remaining adequately decentralised for these differences in what should be encouraged to be brought out. One way to achieve this is to integrate GenAI policy into teaching itself, rather than just stating it in a document on canvas.

Recommendation 3: Clarify departmental policy and integrate policy into teaching so that it is effectively communicated to students and staff.

Recommendation 4: Review the use of AI disclosure forms.

Integration of GenAI into Teaching

52. In our survey, few students reported that ChatGPT Edu was well integrated into their course. On a scale of 1-to-5, where 1 represented poor integration of ChatGPT Edu and 5 represented excellent integration, respondents gave an average of 2.4. This fell to 1.7 among undergraduates and 2.3 among postgraduate taught students. Only among postgraduate research students did the score rise above the neutral point of 3, getting to 3.4. Among divisions, no division had a score greater than 3, though medical sciences scored precisely 3, though this was on a small sample size. Humanities scored the lowest on 2.1.

53. Intriguingly, students who reported feeling that ChatGPT Edu was well integrated into their course also reported that they felt more confident in their work after using ChatGPT Edu, though the correlation is not perfectly clear (Figure 4). If this were true it would fit with the literature that suggests that a more thought-out approach to GenAI in education helps is a useful way to improve educational outcomes by acting on the moderators between GenAI use and outcomes such as cognitive effort.

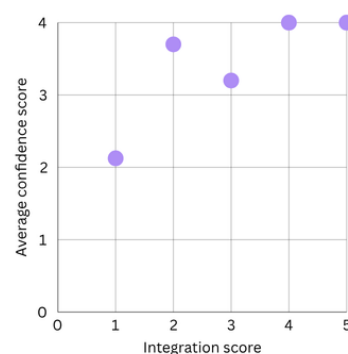


Figure 4: Integration of ChatGPT Edu into teaching, vs average confidence after using ChatGPT Edu

54. Further integrating GenAI into teaching could set the tone for what type of GenAI use is accepted and promoted, encourage teaching staff to build their awareness of policy and help students to develop judgement about the effective and ethical use of GenAI in their learning.
55. This could mean, for one, very clearly stating what a department expects of students with regard to GenAI but also adapting teaching and formative assessment to encourage students to use GenAI appropriately and helping them to use it more appropriately and effectively for their learning and skill development.
56. The main challenge of such integration is the amount of teaching at Oxford is delivered through tutorials which are organised independently by the tutor without very significant

control from the department or even college. As such, to encourage tutors to integrate GenAI into their teaching would require a sustained effort from departments and colleges. If this was achieved, however, it could ensure that policy on GenAI use is still subject-specific, while still giving students the confidence to know whether they have crossed a line.

57. Departments could also create custom GPTs to structure or direct students use of ChatGPT Edu during their learning, as well as ensuring the accuracy of LLM responses. If possible, it could ideally even point out inappropriate use of GenAI, which would reassure students that they are not crossing lines. This would align with the conclusions of the Harvard study, which suggested that custom GenAI tools improve learning outcomes with GenAI.

Recommendation 5: Integrate GenAI into teaching, perhaps using custom GenAI tools, to direct students towards ethical and effective use of GenAI, as well as help students to develop judgement with GenAI tools.

Summative Assessment

58. When it comes to the content of new, clearer policies, assessment policy is probably the most important. This topic was covered extensively in the SU's report last year on undergraduate assessment, but many of the same points recurred this year.
59. For one thing, students are concerned about the shift back towards in-person exams because of GenAI. Students continue to see good reason to use a mixture of assessment methods and conditions to reliably assess different competencies. The rapid shifts, especially to hand written exams, were raised in our focus groups as a particular cause for concern for students who have not spent the rest of their time in higher education being assessed under such conditions. The rapid shifts also contribute to a sense that all students are being punished because some of their peers have used GenAI inappropriately.
60. Students think that the university should be pragmatic on assessment. It is impossible to remove GenAI completely from assessment – students use GenAI to prepare for in-person exams, whether by summarising readings or creating revision materials. Students may or may not accurately declare GenAI use for open-book work. Whether a student used GenAI for open-book work can be hard to prove. Students know this so think the university should encourage students not to use GenAI inappropriately by creating teaching and assessment methods that shape how students behave.
61. Assessment design plays a crucial role in shaping how students approach their learning, what skills they cultivate and how they prepare for assessment (Walker 2021). Including some in-person exams makes sense, because it encourages students not to teach themselves to be dependent on GenAI. On the other hand, there remain benefits to other forms of assessment in terms of what they assess and how reliably they assess different students. Moreover, by including methods like vivas, students can be tested for genuine understanding of material and the ability to put together an argument under closed-book conditions, acting as a test for the validity of preceding open-book work. Departments must think about what competencies they are assessing and teaching, then think about what

different assessment encourage and test for. Once again, thinking about the effect of assessment on learning can help to change the tone on GenAI use.

Recommendation 6: Use a diversity of assessment methods, considering how different assessment methods promote learning, as well as their collective reliability.

Privacy and ChatGPT Edu

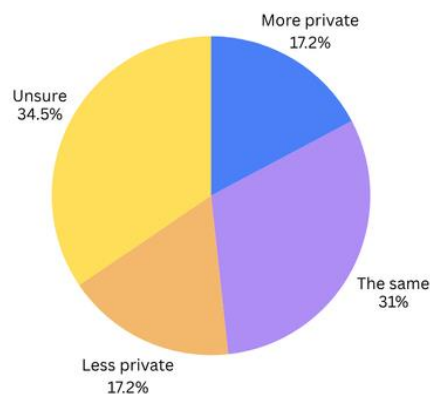


Figure 5: 'Do you feel as though your conversations with ChatGPT Edu are more or less private than with other LLM interfaces?'

62. One final issue around the university's approach to GenAI is the question of privacy. The university purchased ChatGPT Edu licences to ensure data security by limiting OpenAI's access to user data, which could otherwise have been shared by students using less secure personal accounts. Yet, there is anecdotal evidence of students not trusting ChatGPT Edu to be genuinely more private.

63. Our survey offers a mixed picture (Figure 5). Most students were not sure, but 48% of students did not think that ChatGPT Edu was more private and a substantial 17.2% thought their conversations were less private with ChatGPT Edu than with other LLM interfaces. This could reflect a general sense of distrust with the university's approach to GenAI.

64. If it is at all possible, it may help to reassure students if the legal agreement with OpenAI could be published. Such transparency could help to reassure students that their privacy will be respected when using ChatGPT Edu.

Recommendation 7: If possible, publish data privacy agreements with OpenAI to increase trust in ChatGPT Edu.

Environmental and Social Consequences of GenAI at University

65. There was no direct question about the environmental or social consequences of GenAI at university in our survey, but there was an open box for comments in which several students chose to raise concerns about GenAI. 3 students mentioned concerns about Intellectual Property rights, 3 students stated that AI is not useful and therefore should not be used, 2 students raised concerns about OpenAI's dealings with the United States Department of Defence, 2 students pointed to non-specific environmental concerns, and 2 student cited social risks of AI. During our focus groups, students also raised concerns about bad social outcomes from GenAI including risks from GenAI encouraging mental illness, generating misinformation and increasing mistrust, reducing cognitive capacity among young people, removing a human element of life and degrading online aesthetics.

66. Several of these issues were raised during our focus groups. Here, a divide appeared between environmental concerns and other social implications of GenAI. Where students felt like their own usage of GenAI contributed to a social problem, such as when their usage of LLMs increases global energy usage, they felt it was their responsibility to use

GenAI ‘responsibly.’ When asked for what constituted ‘responsible’ use, two suggestions came out, non-trivial use and socially beneficial use, with the former being more common. On the other hand, when social problems resulted from the behaviour of others (such as misinformation or mental illness) these were seen as almost entirely a matter for government regulation of AI companies. While students reported feeling ‘guilty’ about the environmental impact of GenAI, no one in our focus groups said they abstained from GenAI use because of environmental or social concerns. One student who didn’t use GenAI said even they saw GenAI as like meat-eating: bad for the environment, but they wouldn’t judge someone for the environmental impact of meat-eating or for using GenAI. Another who studied climate change said, ‘I do feel a bit guilty when using it because I’m like, oh, in the long term, yeah, it’s just another query but it does take an effect on a system that hasn’t been really prepared for this.’

67. In general, students in our focus groups were concerned with the environmental and social consequences of GenAI and did not think the current state of regulation was adequate. However, there was no sense in our focus groups that these impacts meant that the University should cancel ChatGPT Edu subscriptions as a result. Some students in our survey explicitly advocated ending the contract on ethical grounds, but exactly how representative an opinion this is cannot be determined through a survey with such a small sample.

Technical Details

68. Several points were raised, both in our focus groups and in the open comments box of our survey, that reflected specific details of the ChatGPT Edu system.
69. Some students found themselves hitting the usage limits for OpenAI’s most advanced models and noted frustration with this, sometimes leading them to use personal accounts.
70. Many students who do programming tasks find Anthropic’s Claude Code (which the University does not provide) to be more effective and have sometimes used this at personal expense.
71. Some prefer to continue using personal ChatGPT accounts because they used it before they gained access to ChatGPT Edu and wish to retain access to their conversation history.
72. Others prefer NotebookLM (which the University also provides) because users can upload larger numbers of readings or notes and ask the LLM to use them and generate resources using them.

Recommendation 8: If possible, increase usage limits, allow transfer of conversation history from personal ChatGPT accounts to ChatGPT Edu accounts, and secure agreements for access to a diversity of LLM interfaces.

Conclusion

73. This report has presented key insights into the rollout of ChatGPT Edu.
- We found that students are broadly using ChatGPT Edu carefully and want to preserve their academic skills even while they use it.

- b. Some students feel as though they have to use ChatGPT Edu to keep up – this might be concerning if students are feeling forced to use a tool that will harm their learning. On the other hand, if it can be used to improve learning and productivity at the same time, students' reported focus on using it to reduce workload pressures may suggest that ChatGPT Edu could reduce inequalities and difficulties arising from workload.
- c. Students reported that LLMs are using LLMs to improve education skills and improve their English. This could help to reduce inequalities arising from the hidden curriculum or language skills.
- d. Students support staff use of GenAI, especially due to low pay, but it is crucial that any staff use of GenAI retains the distinct advantage of an expert, human educator.
- e. Students are concerned about the environmental and social impact of GenAI, but mainly see this as a matter for regulation.

74. That said, this report has raised some key concerns about GenAI at the University of Oxford, which gave rise to 8 key recommendations.

Recommendation	Reasoning
1. Further investigate whether and why students with SSPs are not using ChatGPT Edu as frequently as other students.	The survey found that students with SSPs use ChatGPT Edu much less than other students, but our sample size was too small to have confidence in the finding.
2. Monitor inequalities between students, especially between those developing and not developing AI skills, as well as between those with and without English as a first language.	Students reported very different levels of engagement with AI skills, potentially opening up inequalities. GenAI has significant potential to help students with English language skills, but LLMs have less data for other languages so the benefit may be cancelled out.
3. Clarify departmental policy and integrate policy into teaching so that it is effectively communicated to students and staff.	Students reported that university policy is vague and applied inconsistently which made them fearful of GenAI use even when they were confident they were within the rules.
4. Review the use of AI disclosure forms.	Students reported that AI disclosure forms felt like an 'admission of guilt' that discouraged GenAI use, even if they were confident they were within the rules.
5. Integrate GenAI into teaching, perhaps using custom GenAI tools, to direct students towards ethical and effective use of GenAI, as well as help students to develop judgement with GenAI tools.	Students reported greater confidence in their work after using ChatGPT Edu when they also reported better integration of ChatGPT Edu into their course. Literature suggests that custom GenAI tools can ensure learning is directed towards effective GenAI use. Custom GenAI tools could also communicate about the boundaries of permitted use and create more confidence in students' ability to follow rules.

6. Use a diversity of assessment methods, considering how different assessment methods promote learning, as well as their collective reliability.	Students should not be encouraged to be reliant on GenAI (Oxford students seem to be careful about this) and assessment must be reliable, but relying on closed-book exams alone appears to punish all students and limits pedagogical benefits.
7. If possible, publish data privacy agreements with OpenAI to increase trust in ChatGPT Edu.	17% of students think that ChatGPT Edu is less private than other LLM interfaces, despite the University's public claim that its agreement with OpenAI protects data privacy.
8. If possible, increase usage limits, allow transfer of conversation history from personal ChatGPT accounts to ChatGPT Edu accounts, and secure agreements for access to a diversity of LLM interfaces.	These specific technical issues were raised by students as barriers to using ChatGPT.

Bibliography

- AI Competency Cent. (AI Competency Centre). n.d.-a. 'Creating Custom GPTs to Support Teaching and Learning'. Accessed 8 April 2026. <https://oerc.ox.ac.uk/ai-centre/ai-in-education-case-studies/creating-custom-gpts-to-support-teaching-and-learning>.
- (AI Competency Centre). n.d.-b. 'Support for Neurodiverse Humanities Students'. Accessed 8 April 2026. <https://oerc.ox.ac.uk/ai-centre/ai-in-education-case-studies/support-for-neurodiverse-humanities-students>.
- Ananya. 2024. 'AI Image Generators Often Give Racist and Sexist Results: Can They Be Fixed?' *Nature* 627 (8005): 722–25. <https://doi.org/10.1038/d41586-024-00674-9>.
- Ariz. State Univ. (Arizona State University). n.d. 'CreateAI Builder Resources'. Accessed 12 April 2026. <https://ai.asu.edu/technical-foundation/createai-builder>.
- Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakçı, and Rei Mariman. 2024. 'Generative AI Can Harm Learning'. Preprint, SSRN. <https://doi.org/10.2139/ssrn.4895486>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? '. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, March 3, 610–23. <https://doi.org/10.1145/3442188.3445922>.
- Chapman, Rosalie. 2024. *Welfare Survey Analysis*. Oxford SU. <https://www.oxfordsu.org/pageassets/representation/welfaresurveyanalysis/SU-Welfare-Survey-Analysis.pdf>.
- Cheng, Myra, Cinoo Lee, Pranav Khadpe, Sunny Yu, Dyllan Han, and Dan Jurafsky. 2026. 'Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence'. *Science* 391 (6792): eaec8352. <https://doi.org/10.1126/science.aec8352>.
- Deng, Ruiqi, Maoli Jiang, Xinlu Yu, Yuyan Lu, and Shasha Liu. 2025. 'Does ChatGPT Enhance Student Learning? A Systematic Review and Meta-Analysis of Experimental Studies'. *Computers & Education* 227 (April): 105224. <https://doi.org/10.1016/j.compedu.2024.105224>.
- Dennison, Deepak Varuvel. 2025. 'What AI Doesn't Know: We Could Be Creating a Global "Knowledge Collapse"'. News. *The Guardian*, November 18. <https://www.theguardian.com/news/2025/nov/18/what-ai-doesnt-know-global-knowledge-collapse>.
- Edsights. 2024. *How AI Enables More Human Connections for SNHU Academic Advisors. A Randomized Controlled Trial on the Impact of EdSights' AI Framework on Persistence and Advisor Efficiency*.
- Essel, Harry Barton, Dimitrios Vlachopoulos, Albert Benjamin Essuman, and John Opuni Amankwa. 2024. 'ChatGPT Effects on Cognitive Skills of Undergraduate Students: Receiving Instant Responses from AI-Based Conversational Large Language Models (LLMs)'. *Computers and Education: Artificial Intelligence* 6 (June): 100198. <https://doi.org/10.1016/j.caeai.2023.100198>.
- Hall, Rachel. 2025. 'ChatGPT-5 Offers Dangerous Advice to Mentally Ill People, Psychologists Warn'. Technology. *The Guardian*, November 30. <https://www.theguardian.com/technology/2025/nov/30/chatgpt-dangerous-advice-mentally-ill-psychologists-openai>.
- Harkins-Brown, Andrea R., Linda Z. Carling, and David C. Peloff. 2025. 'Artificial Intelligence in Special Education'. *Encyclopedia* 5 (1): 11. <https://doi.org/10.3390/encyclopedia5010011>.
- Henau, Joris de. 2024. *The Impact of Academic Skills Development on Student Outcomes in Higher Education - Literature Review*. Oxford: Centre for Teaching and learning. https://wwwctl.ox.ac.uk/sites/default/files/ctl/documents/media/academic_skills_literature_review.pdf.
- Hougie, Jacob. 2025. *Undergraduate Assessment at Oxford*. Oxford SU.

- Kalai, Adam Tauman, Ofir Nachum, Santosh S. Vempala, and Edwin Zhang. 2025. 'Why Language Models Hallucinate'. Version 1. Preprint, arXiv. <https://doi.org/10.48550/ARXIV.2509.04664>.
- Kestin, Greg, Kelly Miller, Anna Kales, Timothy Milbourne, and Gregorio Ponti. 2025. 'AI Tutoring Outperforms In-Class Active Learning: An RCT Introducing a Novel Research-Based Design in an Authentic Educational Setting'. *Scientific Reports* 15 (1): 17458. <https://doi.org/10.1038/s41598-025-97652-6>.
- Li, Pei-Hsin. 2024. *Student Experience in HEIs: What Does the Evidence Say about the Ethnicity Awarding Gap?* The Diversity of Student Experience Research Project. Centre for Teaching and Learning.
- Lukes, Dominik. n.d. *How to Evaluate Suitability of AI for Academic Tasks*: (Oxford).
- Mountford-Zimdars, A., D. Sabri, J. Moore, J. Sanders, S. Jones, and L. Higham. 2015. *Causes of Differences in Student Outcomes*. HEFCE. <https://research.manchester.ac.uk/en/publications/causes-of-differences-in-student-outcomes-hefce>.
- Murgia, Madhumita. 2023. 'Generative AI Exists Because of the Transformer'. September 12. <https://ig.ft.com/generative-ai/>.
- Nguyen, Andy, Ha Ngan Ngo, Yvonne Hong, Belle Dang, and Bich-Phuong Thi Nguyen. 2023. 'Ethical Principles for Artificial Intelligence in Education'. *Education and Information Technologies* 28 (4): 4221–41. <https://doi.org/10.1007/s10639-022-11316-w>.
- Pava, Juan N., Caroline Meinhardt, Haifa Badi Uz Zaman, et al. 2025. *Mind the (Language) Gap: Mapping the Challenges of LLM Development in Low-Resource Language Contexts*. Stanford Institute for Human-Centred AI. <https://hai-production.s3.amazonaws.com/files/hai-taf-pretoria-white-paper-mind-the-language-gap.pdf>.
- Qian, Yufeng. 2025. 'Pedagogical Applications of Generative AI in Higher Education: A Systematic Review of the Field'. *TechTrends* 69 (5): 1105–20. <https://doi.org/10.1007/s11528-025-01100-1>.
- Rasovic, Ilija, and Evelyn Svingen. 2025. 'Empowering Reflection: Possible Leveraging of Generative Ai to Enhance Personal Tutoring and Student Support in Higher Education'. *Education in Practice* 6 (1).
- Reitmeier, Lea, and Sylvan Lutz. 2025. 'What Direct Risks Does AI Pose to the Climate and Environment?' *Grantham Research Institute on Climate Change and the Environment*, September 12. <https://www.lse.ac.uk/granthaminstitute/explainers/what-direct-risks-does-ai-pose-to-the-climate-and-environment/>.
- Stadler, Matthias, Maria Bannert, and Michael Sailer. 2024. 'Cognitive Ease at a Cost: LLMs Reduce Mental Effort but Compromise Depth in Student Scientific Inquiry'. *Computers in Human Behavior* 160 (November): 108386. <https://doi.org/10.1016/j.chb.2024.108386>.
- Stephenson, Rose, and Charlotte Armstrong. 2026. *Student Generative AI Survey 2026*. No. 199. Higher Education Policy Institute.
- Tao, Yan, Olga Viberg, Ryan S. Baker, and René F. Kizilcec. 2024. 'Cultural Bias and Cultural Alignment of Large Language Models'. *PNAS Nexus* 3 (9): pgae346. <https://doi.org/10.1093/pnasnexus/pgae346>.
- Univ. Edinb. Inf. Serv. (University of Edinburgh Information Services). 2024. 'ELM'. October 8. <https://information-services.ed.ac.uk/computing/comms-and-collab/elm>.
- Universities UK and National Union of Students. 2019. *Black, Asian and Minority Ethnic Student Attainment at UK Universities: #CLOSINGTHEGAP*. London: Universities UK. <https://www.universitiesuk.ac.uk/sites/default/files/field/downloads/2021-07/bame-student-attainment.pdf>.
- University of Oxford. 2025. 'AI Use in Summative Assessment'. <https://governance.admin.ox.ac.uk/education-committee/policies/ai-use-in-summative-assessment>.

- Voinea, Cristina, Sebastian Porsdam Mann, Julian Savulescu, and Brian D. Earp. 2026. 'The Calculator Analogy: Epistemic Virtues for Using LLMs'. *Technology in Society* 85 (June): 103198. <https://doi.org/10.1016/j.techsoc.2025.103198>.
- Walker, Ruth. 2021. *Awarding Gaps and Assessment: A Briefing Paper for EAC*. No. 1. Foundations in Assessment. Cambridge Centre for Teaching and Learning.
- Wilson, Thomas D. 2025. 'The Development of Policies on Generative Artificial Intelligence in UK Universities'. *IFLA Journal* 51 (3): 722–34. <https://doi.org/10.1177/03400352251333796>.

AI Declaration

Generative AI was used in the preparation of this report in the following ways:

- ChatGPT Edu was used to fix Excel formulas during data analysis.
- ChatGPT Edu was used at times to find additional source material.
- ChatGPT Edu and Claude reviewed already written content for accuracy and readability.
- ChatGPT Edu generated the sample prompts used in the survey, using detailed instructions.
- Google AI studio was used to transcribe the recording of the focus group, which was verified when analysing the transcript.