# Royal Holloway's AI toolkit: redefining ethical use and collaboration in higher education



Universities have grappled with the dual challenge of fear and fascination towards generative artificial intelligence (GenAI) over the past two years, seeking ways to integrate this evolving technology meaningfully into education. While academic staff often debate the risks and potentials of AI, most students simply want clarity on what constitutes acceptable use that supports learning without risking academic integrity. In response to this complex landscape, Royal Holloway, University of London, has developed the AI Collaboration Toolkit, a practical, student-focused framework designed to treat AI not as a threat, but as a literacy to be taught and responsibly integrated into academic work.

The toolkit rests on a foundational promise: AI exists to augment students' efforts rather than replace their intellectual contribution. Drawing on the late Clayton Christensen’s Jobs-to-Be-Done theory, it encourages students to define the specific task they want AI to assist with—be it brainstorming, structuring, revising, or translating—and to “hire” the appropriate AI tool accordingly, while retaining ultimate accountability for originality and judgement. This approach goes beyond the simplistic binary of “ban it or embrace it” and fosters a disciplined, purposeful use of AI, aligning tools to tasks with human decision-making at the core.

Central to the AI Collaboration Toolkit is a Typology of Student AI Use, a concise matrix categorising common tasks alongside typical AI tools, the roles played by both human and machine, patterns of interaction, and a Red-Amber-Green (RAG) risk rating that signals the acceptability of AI use. For instance, low-risk activities like brainstorming ideas or obtaining formative feedback fall into the green zone, while drafting entire assignments using AI is categorised as high-risk and red. Intermediate, amber-rated activities such as paraphrasing complex texts or auto-generating citations require careful verification and declarations. This typology helps delineate clear boundaries for students and staff, ensuring transparency and fairness in assessment.

Equity and inclusivity are vital considerations within the toolkit. It acknowledges that some students, particularly those who speak English as a second language or have limited access to one-to-one academic support due to caregiving or work commitments, can benefit from accessible AI supports. Fluency improvement and translation of personal ideas are recognised as low-risk with appropriate critical review and clear authorship. The toolkit also encourages AI-assisted planning, structuring, and argument clarification as ways to reduce barriers and foster genuine learning. By defining what “good help” looks like, it offers uncertain students a dignified, supportive framework to develop their work and academic confidence.

To reinforce ethical literacy, the toolkit integrates two reflective practices: an AI Reflection Journal and a Final Self-Check. The journal prompts students to document the task undertaken, the AI’s contribution, lessons learned, and the steps taken to ensure academic integrity. This reflection demystifies AI’s role, transforming it from a black box into a tool subject to critical interrogation. The self-check asks questions such as whether AI was used solely to support thinking, whether the student can explain their work if challenged, and if all citations have been verified, nudging learners toward deeper understanding rather than superficial compliance.

The toolkit also addresses the often confusing terrain of AI-related misconduct through straightforward, plain-English definitions distinguishing AI commissioning (outsourcing authorship), falsification (fabricating data or citations), unauthorised use, and plagiarism (passing off AI-generated content as one’s own). Case vignettes illustrate these breaches in context, helping staff make consistent decisions and students understand the line between acceptable support and academic dishonesty.

Royal Holloway’s approach aligns with broader educational efforts to responsibly incorporate GenAI. For example, University College Cork’s AI toolkit focuses on helping staff and students understand the ethical dimensions of GenAI use in teaching and learning, while LexisNexis offers a curriculum toolkit aimed at developing students’ critical thinking and ethical awareness around AI in academic research. Meanwhile, the University of South Carolina provides interactive modules encouraging ethical AI use in group settings. These initiatives collectively reflect a growing consensus that integrating AI literacy into education should balance opportunity with safeguards, equipping both educators and students for the complexities of GenAI.

Beyond practical guidance, recent academic frameworks such as the AI Assessment Scale (AIAS) and its advanced iteration, the Comprehensive AI Assessment Framework (CAIAF), offer structured approaches to determining when and how AI use supports learning objectives. These frameworks advocate transparent decision-making and ethical standards that echo Royal Holloway’s emphasis on purposeful AI collaboration, prioritising human judgement and maintaining assessment integrity.

In practical terms, the AI Collaboration Toolkit envisions scenarios such as that of Amara, an international first-year student balancing part-time work. Amara employs AI judiciously: she uses it to brainstorm case study angles, drafts her own outline, leverages AI for formative feedback on structure and clarity, maintains an AI Reflection Journal, verifies each reference, and completes a self-check before submission with a concise declaration of AI use. When called upon to defend her argument orally, she can do so confidently, demonstrating that the AI supported—not substituted—her learning.

This model highlights the potential shift universities need: moving from fear and policing of AI usage to partnership and capability-building, where rules are clear, transparency is standard, and assessment focuses on genuine learning. Rather than pretending AI is absent or relinquishing authorship to machines, the goal is to teach students to collaborate with AI ethically and inclusively, upholding trust and critical thinking.

For institutions contemplating similar steps, Royal Holloway recommends three initial moves: establish a shared language for AI-enabled tasks, encourage reflective practices that make thinking visible, and permit transparent declarations of AI use. Coupled with alignment of assessment criteria to valued capabilities and joint staff-student training, these steps can nurture integrity, belonging, and improved academic work in the AI era.

Lucy Gill-Simmen and Will Shüler of Royal Holloway’s Schools of Business and Management and Performing and Digital Arts respectively, articulate a vision of AI as a partner in learning—one that demands human oversight but offers new pathways to fairness and creativity in higher education.

### 📌 Reference Map:

* Paragraph 1 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated), [[2]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated)
* Paragraph 2 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated), [[2]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated)
* Paragraph 3 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated)
* Paragraph 4 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated), [[2]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated)
* Paragraph 5 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated)
* Paragraph 6 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated)
* Paragraph 7 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated)
* Paragraph 8 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated), [[3]](https://www.ucc.ie/en/ethical-use-of-generative-ai-toolkit/), [[4]](https://perspectives.lexisnexis.com/ai-curriculum-toolkit/index.html), [[5]](https://guides.library.sc.edu/info-toolkit/genai)
* Paragraph 9 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated), [[6]](https://arxiv.org/abs/2312.07086), [[7]](https://arxiv.org/abs/2407.16887)
* Paragraph 10 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated)
* Paragraph 11 – [[1]](https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated)

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated> - Please view link - unable to able to access data
2. <https://www.timeshighereducation.com/campus/turn-ai-magic-box-tool-be-interrogated> - This article discusses the development of the AI Collaboration Toolkit by Royal Holloway, University of London, aiming to integrate generative AI into education responsibly. The toolkit provides a framework for students to use AI as a collaborator, with clear ethical guidelines and boundaries, promoting critical thinking and originality. It includes a Typology of Student AI Use, categorising tasks based on risk levels, and encourages reflective practices to ensure academic integrity. The toolkit also defines various forms of AI misconduct to guide students and staff in ethical AI usage.
3. <https://www.ucc.ie/en/ethical-use-of-generative-ai-toolkit/> - University College Cork offers a toolkit designed to assist staff in integrating Generative Artificial Intelligence (GenAI) into teaching and learning practices responsibly. The resource provides a comprehensive understanding of GenAI, its applications, and challenges in education. Developed as part of the (AI)²ed project, it aims to educate both staff and students on the ethical dimensions of GenAI, promoting responsible use in academic settings.
4. <https://perspectives.lexisnexis.com/ai-curriculum-toolkit/index.html> - LexisNexis has developed a GenAI Curriculum Toolkit to help educators teach students about the responsible use of AI in academic research. The toolkit includes lesson plans, practical exercises, and discussion prompts focused on AI ethics, aiming to prepare students to use AI tools thoughtfully and transparently. It addresses the importance of critical thinking, problem-solving, and awareness of the ethical implications of AI, aligning with the World Economic Forum's insights on digital and AI literacy.
5. <https://guides.library.sc.edu/info-toolkit/genai> - The University of South Carolina provides an interactive module to help students identify appropriate scenarios for using Generative AI in academic settings and understand associated risks. The module includes a 'Using Generative AI' worksheet for small group activities, adapted for USC students and faculty, engaging users with ethical methods of using GenAI in education. Additional resources, such as a beginner's guide to Generative AI and guides on ethical AI use, are also available.
6. <https://arxiv.org/abs/2312.07086> - This paper introduces the AI Assessment Scale (AIAS), a framework for the ethical integration of Generative AI in educational assessments. The AIAS empowers educators to select appropriate levels of GenAI usage based on desired learning outcomes, offering clarity and transparency for students and educators. It provides a nuanced approach that embraces GenAI's opportunities while recognising instances where such tools may not be pedagogically appropriate, aiming to address current uncertainties regarding GenAI in education.
7. <https://arxiv.org/abs/2407.16887> - The Comprehensive AI Assessment Framework (CAIAF) is an evolved version of the AI Assessment Scale (AIAS), targeted toward the ethical integration of AI into educational assessments. Incorporating stringent ethical guidelines and clear distinctions based on educational levels, the framework offers an intuitive approach to integrating AI tools into education. Developed through a thorough literature review and practical insights, CAIAF aims to ensure better learning outcomes, uphold academic integrity, and promote responsible use of AI in modern educational practices.