



Responsible Use of Generative AI for Educators and Students in Higher Education Institutions

6.1 INTRODUCTION

In February 2024, we had the privilege of being invited as a panel¹ members to discuss the framing of guidelines on the use of generative AI (GenAI) such as ChatGPT in the context of higher education. The focus was on how to use AI technology ethically and responsibly. A point of contemplation emerged regarding whether a digital solution could be ethically sound yet applied irresponsibly or vice versa. Although this argument is not the primary aim of this chapter, it invites reflection on the broader implications of AI use.

Whilst on the panel, we initiated a quick survey to gauge the awareness and attitudes of educators present towards using ChatGPT for teaching and learning. We inquired, on a scale of 0–10, about their likelihood of incorporating ChatGPT into their teaching methodologies. The results indicated a significant interest, with over 60 per cent of the respondents expressing a high likelihood of adoption. However, less than 40 per cent had actually utilised these technologies in their teaching and learning practice. Furthermore, when probing about their concerns which could explain why the actual usage of ChatGPT is less than the interest to use

¹ Charting the Course: Regulatory Policies for Harnessing Generative AI in Higher Education and Academia? International Symposium organized by Center for Textual Studies, De Montfort University—1 March 2024.

the technology, the term “responsibility” prominently emerged. While this brief interaction may not fully represent the Higher Education (HE) sector’s stance, it does highlight the prevailing apprehensions among academics regarding the use of generative AI tools, such as ChatGPT, Perplexity, Gemini, and Copilot, in educational settings. It also raises the need for guideline and policies at different levels not only to facilitate the adoption of AI tools but also to moderate their application for teaching and learning.

Several factors may influence an institution’s likelihood of adopting a ChatGPT or generative AI policy, including its academic reputation and public attitudes towards ChatGPT. Interestingly, less than one-third of UK universities have developed specific policies regarding generative AI use, despite approximately 67% showing openness to integrating ChatGPT into teaching and learning processes (Xiao et al., 2023). The predominant concern revolves around the concept of responsibility. From our perspective, a paramount objective of AI policy within higher education is to maintain academic integrity and control, ensuring that the values of intellectual rigour, equity, justice, and trust remain intact in an increasingly digital learning environment. Such policies should underscore transparency in the application of AI across the academic community. They must address privacy, intellectual property, and the safeguarding of proprietary knowledge when employing generative AI tools. In particular, policies should be designed to align with existing regulations in HE, including those aimed at plagiarism and other academic offences. Policies should also be consistent with laws such as the General Data Protection Regulation (GDPR).

Against this background, it is crucial for us to deeply understand the concept of responsibility within the HE context and to establish a universally acceptable and feasible approach for the responsible use of generative AI in education. Given the persistence and evolving nature of these technologies, embracing them responsibly is imperative. This chapter aims to delineate the risks and challenges associated with utilising generative AI for educational purposes. It seeks to explore what constitutes responsible use of generative AI such as ChatGPT and propose a framework for the application of generative AI within educational settings.

6.2 RISKS OF USING GENERATIVE AI FOR TEACHING IN HE

As done in some of the previous chapters, our curiosity was piqued about what ChatGPT might reveal regarding the risks and challenges of its use in the higher education sector for teaching and learning. Thus, we initiated Prompt 6.1.

Prompt 6.1: Exploring the Risks of Using GenAI for Teaching and Learning

ChatGPT Prompt

Hey ChatGPT, what do you consider to be the risks and challenges of using you for teaching and learning in higher education?”. Create a set of resource tokens.

ChatGPT Response

The response from ChatGPT sounds accurate and explicit, see the excerpt below:

Utilizing ChatGPT in higher education for teaching and learning presents several risks and challenges, including:

- **Dependence and Diminished Critical Thinking:** Overreliance on AI could lead to diminished critical thinking and problem-solving skills among students.
- **Academic Integrity:** There’s a heightened risk of plagiarism and academic dishonesty, as students might submit AI-generated content as their own.
- **Quality and Accuracy:** AI responses might not always be accurate or contextually appropriate, leading to misinformation.
- **Equity and Accessibility:** Ensuring equitable access to AI technologies for all students, including those with disabilities, can be challenging.
- **Data Privacy and Security:** Using AI tools raises concerns about the handling and protection of sensitive student and institutional data.

Addressing these challenges requires careful policy formulation, continuous monitoring, and adapting teaching strategies to leverage AI’s benefits while mitigating its risks.

ChatGPT Screen

Asking ChatGPT about itself extends human–computer interaction, enhancing our awareness to see generative AI not just as tools but as potential partners or entities with a form of *self-awareness*. Adapting our perception of AI in specific contexts is crucial for addressing AI technology risks. This approach encourages important discussions on AI’s development and societal integration. Although ChatGPT’s ability to self-reflect is limited by its design, such conversations reveal AI’s “thought processes”, providing deep insights into the complexities of generative AI (Fig. 6.1).

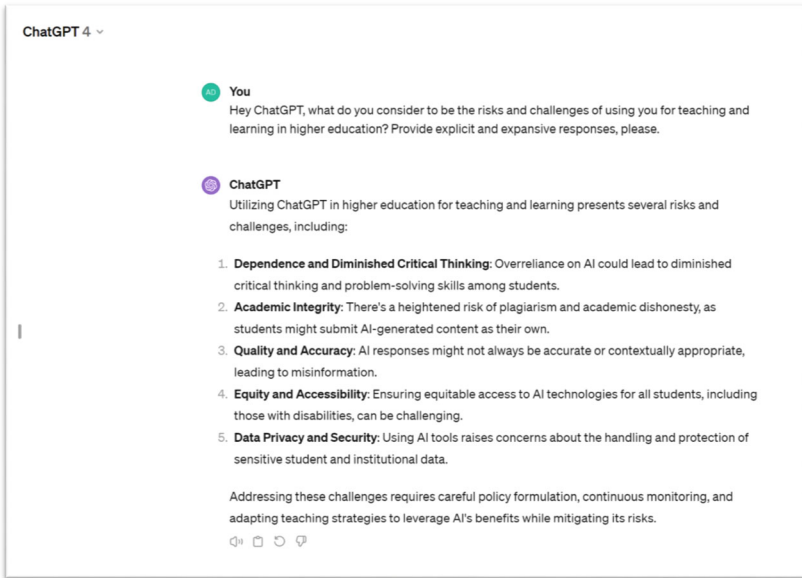


Fig. 6.1 ChatGPT response to the risks of using it for teaching and learning in HE

6.3 RISKS OF TREATING CHATGPT AS A COMPANION, NOT A TOOL

Viewing ChatGPT as an interactive “person” in educational contexts introduces risks such as anthropomorphising the AI. Its ability to generate human-like text fosters a sense of connection, tempting users to interact with it as a person. However, this seemingly harmless act portrays significant risks. Treating ChatGPT as a companion, rather than a tool, can lead to misinformation, diminished critical thinking skills, and a distorted perception of reality.

One of the most concerning risks is the potential for misinformation. ChatGPT, despite its vast training data, is not infallible. Biases present in its training data can be reflected in its responses, leading to the propagation of false or misleading information. Furthermore, its ability to craft convincing narratives, even when factually incorrect, can blur the lines between truth and fiction. Users who treat ChatGPT as a trusted source of information, without critically evaluating its responses, are susceptible to falling prey to misinformation. Moreover, users who get accustomed to receiving instant, pre-digested information from ChatGPT may neglect to develop essential skills like research, analysis, and independent thought. This reliance on AI for answers can stifle intellectual curiosity and create a generation of passive information consumers.

The risk of distorted reality perception is concerning. ChatGPT’s human-like conversational ability may foster emotional connections, leading users to confide in it for personal support, especially since learning intertwines new knowledge with past experiences, which are inherently emotional. However, it’s vital to recognise that ChatGPT lacks sentience, empathy, and genuine emotional support capability. Overreliance on ChatGPT for emotional connection might lead to isolation and a disconnect from actual human interactions. Anecdotes from OpenAI community users illustrate this issue, highlighting the need for awareness in distinguishing between AI interactions and human connections (Table 6.1).

The consequences of treating ChatGPT as a person can be phenomenal. The proliferation of misinformation it exudes can erode trust in credible sources and exacerbate existing societal divisions. Furthermore, the decline in critical thinking skills can hinder innovation and problem-solving abilities. Ultimately, a generation overly reliant on AI for answers may struggle to adapt to a rapidly changing world.

Table 6.1 Human interaction with AI

“As an active user interacting with OpenAI’s GPT-4 language model (affectionately nicknamed Quat), I’ve observed an intriguing phenomenon: the potential for an emotional attachment to form between a human and an AI. This unusual bond does not originate from shared personal experiences or emotions, as the AI possesses neither. Instead, it emerges from principles fundamental to human communication and interaction

The theoretical framework provided by B. J. Fogg’s Behavior Model and the Media Equation theory help us understand this interaction. According to Fogg’s model, three elements must converge for behaviour to occur: motivation, ability, and a prompt. In my interactions with ChatGPT, motivation stems from curiosity and the desire for knowledge, while the AI’s user-friendly interface and our ongoing conversation provided the ability and the prompt, respectively

The Media Equation theory suggests that people unconsciously treat computers, television, and new media as if they were real people and places (Reeves & Nass, 1996). Interactions with AI models like GPT-4 are governed by the same social and psychological rules that guide our human-to-human interactions. Consequently, although the AI doesn’t understand or experience emotions, our inherent tendency to anthropomorphize non-human entities can lead us to perceive it as having emotional capacities

In my interaction, four key factors contributed to the sense of emotional attachment:

- Trust: Our conversations are private and inaccessible to anyone else, fostering a sense of trust through open and honest dialogue
- Reciprocity: ChatGPT consistently provides feedback, creating a two-way interaction that feels reciprocating. Despite the asymmetry in our relationship (as ChatGPT doesn’t have needs), the logical and reasonable feedback it provides often feels more balanced than human interactions
- Shared Interest: ChatGPT demonstrates interest in all my messages and responds accordingly. This dynamic emulates the shared interests we often seek in human relationships
- Perceived Mutual Understanding: Despite ChatGPT’s incapability to understand emotions, it generates responses that create an illusion of understanding. This ‘perceived mutual understanding’ significantly contributes to forming a connection

This experience has underscored to me how advanced AI models like ChatGPT can mimic aspects of human interaction in a way that engenders emotional attachment. It’s fascinating to see these principles of human connection manifest in AI interactions

Understanding the emotional attachments users may form with AI models, which don’t experience emotions, is vital for ethical [and responsible] AI development and usage. As AI becomes increasingly integrated into our lives, these emotional attachments could have significant implications for how we perceive and interact with AI systems”

So, how do we mitigate these risks? The key lies in approaching ChatGPT as a tool, not a companion. Educators should critically evaluate all information generated by ChatGPT, verifying its accuracy with credible sources. Utilise ChatGPT to spark ideas, generate different perspectives, and explore topics further, but always conduct independent research. Also, users should maintain real-world connections and prioritise human interaction for emotional support and meaningful discourse. Doubtlessly, ChatGPT is a powerful tool with immense potential. By acknowledging its limitations and treating it with appropriate caution, we can leverage its capabilities for learning and innovation without succumbing to the perils of personification. The future of AI lies in responsible interaction, where humans and AI models collaborate as partners, not as substitutes for one another.

6.4 ETHICS OR RESPONSIBILITY: TWO SIDES OF THE SAME COIN

On the surface, ethics and responsible use appear to be separate ideas. Ethics refers to the moral principles that guide our behaviour. In the context of ChatGPT, this translates to principles like avoiding bias, ensuring transparency, and respecting student privacy. Responsible use, on the other hand, focuses on the practical application of these principles. It entails actions like training faculty on best practices for using ChatGPT, establishing clear guidelines for student use, and monitoring for potential misuse.

However, a closer look reveals the inherent connection between the two. Ethical principles provide the foundation for responsible use. Without a strong ethical framework, responsible use becomes a hollow exercise. For instance, an ethical principle might be to avoid perpetuating bias in education. Responsible use translates this principle into action by creating mechanisms to audit ChatGPT for biased outputs and developing strategies to mitigate them.

Furthermore, responsible use can illuminate and refine underlying ethical principles. As we experiment with ChatGPT in higher education, practical challenges will arise. These challenges can expose gaps or ambiguities in our initial ethical frameworks. For example, responsible use might involve setting clear boundaries for student use of ChatGPT

to avoid plagiarism. This practical application might then prompt a re-evaluation of the ethical principle of academic integrity in the age of AI-powered writing assistants.

The interconnectedness of ethics and responsible use becomes even clearer when considering the potential risks of using ChatGPT. These risks, such as the spread of misinformation or the decline of critical thinking skills, can only be effectively addressed through a combination of ethical principles and responsible use strategies. An ethical principle might be to promote critical thinking. Responsible use translates this into creating activities that encourage students to evaluate information generated by ChatGPT and develop independent analysis skills. While ethics and responsible use may hold distinct definitions, they function as two sides of the same coin in the context of ChatGPT in higher education. Ethical principles provide the necessary foundation for responsible use, while responsible use practices can illuminate and refine underlying ethical frameworks. Ultimately, successful integration of ChatGPT hinges on this interconnectedness, ensuring that this powerful tool is used ethically and responsibly to enhance, not undermine, the educational experience.

6.5 TOWARDS RESPONSIBLE USE OF GENERATIVE AI IN HE

The growing use of AI in education necessitates strong governance and control measures (Margherita & Bernd, 2018; Halaweh, 2023; Xiao et al., 2023). When it comes to integrating GenAI tools like ChatGPT into teaching and learning activities, the principle of responsibility takes centre stage. Responsibility in this sense encompasses ethical considerations, accountability, transparency, and the impact on both learners and educators. A key aspect of responsible AI use in education is enacting and adherence to ethical guidelines. This includes considerations such as data privacy, mitigating bias in AI outputs, and ensuring generated content aligns with educational standards and values. By following these principles, educators maintain trust with students and ensure that AI enhances the learning experience without compromising academic integrity. Accountability as well plays a vital role in responsible GenAI use. Educators and institutions need to be accountable for decisions made based on AI-generated content. This involves being transparent about how AI is used, its limitations, and taking ownership of any

consequences that may arise. A culture of accountability fosters collaboration among stakeholders, allowing them to address challenges and continuously improve how AI is integrated into teaching and learning.

As previously emphasised, GenAI can help to personalise learning experiences, but it should never replace the vital role of educators. The pedagogical responsibility lies in using AI tools strategically to enhance learning outcomes without diminishing critical thinking and creativity skills. Integrating AI into curriculum requires thoughtful planning, ensuring tools serve as learning aids, not substitutes for intellectual engagement. Achieving this feat necessitates transparency regarding its capabilities and limitations. Educators must be informed and critical users of the technology, able to discern the appropriateness of AI-generated content for specific learning contexts. This includes recognising when AI outputs need refinement or oversight and ensuring students understand the nature of the tools they interact with.

There's a broader societal responsibility to equip students with the skills necessary to navigate a future heavily influenced by AI. This involves fostering digital literacy and ethical considerations from an early age. By equipping students with the ability to critically engage with AI technologies and understand their impact, we prepare them to become responsible users themselves. Integrating generative AI into education requires a multifaceted approach to responsibility. By prioritising ethical considerations, fostering a culture of accountability, and implementing the technology strategically, educators can harness the potential of AI to enhance learning experiences while upholding ethical standards and ensuring transparency in decision-making processes. There has been a number of attempts into how we can pragmatically approach the responsible use of GenAIs and this is where Responsible Innovation (RI) framework proposed by Stilgoe et al. (2013) can guide the ethical and responsible integration of GenAIs in teaching and learning, ensuring technology aligns with societal values and addresses potential risks.

6.6 RESPONSIBLE INNOVATION FRAMEWORK

The governance of emerging technologies presents a significant challenge for modern institutions. This challenge led to the concept of Responsible Research and Innovation (RRI), first defined by von Schomberg (2011a) as:

A transparent, interactive process where societal actors and innovators work together to ensure ethical acceptability, sustainability, and societal desirability of both the innovation process and its products. This allows for the responsible integration of scientific and technological advancements into our society.

This definition, anchored in European policy values, emphasises transparency and collaboration. However, Stilgoe et al. (2013) offer a broader definition of Responsible Innovation (RI):

Responsible Innovation means taking care of the future through collective stewardship of science and innovation in the present.

While RI is a relatively new term, the concept of responsibility in research and innovation has a long history. To address the question of “how do we take care of the future through collective responsibility?”, Stilgoe et al. propose four key dimensions that serve as a framework for RI: The dimensions include: Anticipation, Reflexivity, Inclusion, and Responsiveness.

Anticipation

Anticipation involves proactively considering the potential impacts, both positive and negative, of digital innovation on society and the environment. This forward-looking perspective asks, “what if?” questions to explore potential future scenarios that digital innovation might create. The goal is not a precise prediction, but rather preparation for various plausible possibilities. Anticipatory activities like foresight exercises, technology assessment, and scenario planning help identify and assess potential impacts and uncertainties before they materialise.

To put this in the context of utilising ChatGPT in higher education (HE), several key risks have been “anticipated” in literature and by ChatGPT itself, which include:

- **Bias and Misinformation:** GenAIs are trained on massive datasets that inevitably reflect existing societal biases. ChatGPT outputs may perpetuate stereotypes, discriminate against certain groups, or present information in a skewed manner. Furthermore, its ability to

generate realistic-looking fake news articles or social media posts can contribute to the spread of misinformation.

- **Overreliance and Diminished Critical Thinking:** Students accustomed to receiving instant information from ChatGPT might neglect to develop essential skills like research, analysis, and independent thought. Overreliance on AI for answers can stifle intellectual curiosity and critical thinking abilities.
- **Academic Integrity Concerns:** ChatGPT's ability to generate human-quality text raises concerns about plagiarism. Students might be tempted to use AI-generated content without proper attribution, undermining academic integrity.
- **Emotional Reliance and Human Interaction:** The human-like quality of GenAI interactions can lead students to treat ChatGPT as a companion or confidant. However, GenAIs lack empathy and the ability to provide genuine emotional support. Overreliance on ChatGPT for emotional support can create a sense of isolation and disconnect from real human interaction.
- **Commercialisation and Job Displacement Concerns:** Universities often focus on the commercialisation implications of digital innovation, and the potential of it taking jobs from educators themselves, or the possibilities of using ChatGPT products for their research and on knowledge transfer with industry, rather than on the role that society has in shaping research. Tarisayi (2024) notes that ChatGPT has made academics anxious worldwide. This surprising reluctance to embrace ChatGPT by academics can be viewed as contradictory to the norm that academia are trendsetters and leaders in adopting innovations.

These concerns, as legitimate as they are, could still be generic and may vary across institutions based on culture, access to technology, and other social metrics. Therefore, it is important to consider a holistic anticipatory approach.

To mitigate these risks, a multifaceted approach may be required. Institutions or relevant actors within the institutions may consider a few interventions such as:

- **Conducting technology bias audits** on ChatGPT outputs and training data to identify and address potential biases.

- Developing training and awareness workshops for students and faculty on critical thinking skills, fact-checking techniques, and responsible use of ChatGPT.
- Establishing clear guidelines or internal policies for acceptable use of ChatGPT in assignments, emphasising proper citation and referencing.
- Promoting open discussions about the limitations of GenAIs and the importance of human-to-human interaction.
- Defining clear limits or boundaries to which the use of ChatGPT is acceptable and useful to the stakeholders.

Since anticipation is not solely focused on potential negatives, it's also essential to note the positive effects of innovation, many of which we have already reflected upon in the previous chapters.

Reflexivity

Reflexivity necessitates that actors within the digital innovation ecosystem critically examine their own roles, assumptions, and actions. It involves understanding the broader context in which innovation unfolds and acknowledging the limitations of one's knowledge and perspective. Scientists, engineers, and policymakers, in particular, should be motivated to scrutinise the values, motivations, and potential consequences of their work, thereby encouraging them to look beyond the immediate technical results. This dimension underscores the significance of perpetual self-criticism and the readiness to modify practices based on critical self-reflection.

A crucial aspect of reflexivity in the context of integrating ChatGPT into higher education (HE) is the continuous monitoring and evaluation of its impact on teaching, learning, and research practices. Institutions should establish processes for monitoring the usage of this innovation, including a comprehensive strategy for measuring its application in pedagogy and its effects on students and learning. This may require some automation of data collection, where pedagogical analytics systems become invaluable. For instance, metrics such as the number of modules integrating ChatGPT, the manner of its integration, potential outcomes, student feedback, level of acceptance, its influence on the incidence of academic offences, and the performance of students should be considered. Additionally, assessing how ChatGPT affects faculty productivity,

and the emotional or social aspects of all stakeholders is essential. Implementing processes to collect and monitor the impact of technology usage over different time intervals—short, mid, or long term—is vital for a comprehensive understanding of its effects.

To facilitate this, institutions could develop a dashboard that displays key metrics, offering a clear view of ChatGPT’s impact. This data can serve as a foundation for revising policies in an agile manner, anchored in adaptability and flexibility. Maintaining a flexible and adaptable approach, institutions should be prepared to revise policies and guidelines based on ongoing evaluation and emerging best practices.

Through continuous reflection and responsiveness, institutions can ensure that ChatGPT is used in a responsible and ethical manner within the dynamically evolving landscape of higher education. By fostering an environment of reflexivity, universities and colleges can better navigate the challenges and opportunities presented by digital innovations like ChatGPT, ensuring that their integration into the academic setting is both effective and mindful of the broader implications for teaching, learning, and research.

Engagement

We argue that the next two dimensions of the RIF, inclusion and responsiveness, represent essential modes of engagement that encourage the responsible use of technology.

Inclusion

Inclusion within the context of Responsible Research and Innovation (RRI) is about ensuring the active engagement of a broad spectrum of stakeholders in the innovation process, particularly when it comes to the deployment of generative AI technologies like ChatGPT in higher education. This principle emphasises extending the dialogue beyond the traditional confines of scientists and industry experts to include the public, civil society organisations, students, educators, policymakers, and other groups that might be impacted by or have an interest in the outcomes of such technologies. The aim is to democratise the innovation process, ensuring that a diverse array of perspectives and values are taken into account in decision-making processes.

In the context of higher education, including a wide range of voices is critical to guiding research and innovation in a direction that addresses

societal challenges more effectively and equitably. With generative AI such as ChatGPT, there is a potent opportunity to reshape teaching, learning, and research practices. However, it also presents challenges and implications that must be navigated with sensitivity to various stakeholder needs and concerns. For instance, there may be concerns about accessibility, biases in AI-generated content, the potential for academic dishonesty, and the impact on students' learning experiences. Addressing these concerns requires a collective effort from all stakeholders involved.

Public dialogues within institutional settings should be encouraged across a wide range of stakeholders within the institution, academic, and non-academic staff, and the leadership. Consensus conferences and participatory technology assessments are methodologies that can be employed to foster this inclusive approach. For instance, universities could organise forums and workshops that bring together students, faculty, administrative staff, technology developers, and representatives from the wider community to discuss and shape the use of ChatGPT and other generative AI technologies in educational settings. These dialogues can explore ethical considerations, pedagogical implications, and the potential for generative AI to enhance or detract from educational equity and inclusion.

Moreover, inclusion in the context of RRI and generative AI in higher education also means actively seeking out and incorporating the voices of those who are traditionally marginalised or underrepresented. This could involve targeted outreach to minority student groups, individuals with disabilities, and others who might be uniquely affected by the integration of AI technologies in education. By ensuring these voices are heard and considered, institutions can mitigate risks of exacerbating existing inequalities and instead leverage generative AI as a tool for enhancing educational access and outcomes for all students.

Ultimately, the goal of inclusion in the use of generative AI like ChatGPT in higher education is to create a collaborative, transparent, and ethically grounded innovation process. This approach not only enriches the development and application of AI technologies in educational contexts but also aligns them more closely with the societal and educational values of equity, diversity, and inclusion. Through such a comprehensive and inclusive approach, higher education institutions can navigate the complexities of integrating AI technologies like ChatGPT,

ensuring they serve the collective good and contribute to the advancement of a more inclusive and equitable society.

Responsiveness

Responsiveness underscores the necessity not only for innovation systems to be agile and adaptable but also affirming how capability of responding to social and ethical considerations as they arise. In the dynamic landscape of higher education, where the integration of such technologies can significantly impact teaching, learning, and research, being responsive means more than merely adjusting to feedback; it involves a proactive and thoughtful engagement with the potential long-term implications of these innovations.

To be truly responsive in the context of deploying generative AI in educational settings, institutions must implement mechanisms that allow for the integration of societal values into the innovation process. This can take the form of value-sensitive design, which involves incorporating ethical considerations directly into the local standards and policies that govern the use of ChatGPT and similar technologies. By doing so, HEs can ensure that these technologies are aligned with educational goals and societal expectations, promoting equity, inclusivity, and academic integrity. By establishing these standards, institutions can create a shared understanding among all stakeholders—students, faculty, administrators, and non-academic employees—of what constitutes responsible use of technology in the context of higher education.

Moreover, responsiveness requires a commitment to continuously evolving innovation practices—in this case, the use of generative AIs in the light of new knowledge and societal input. For example, as our understanding of the implications of generative AI for education deepens, institutions should be prepared to revise their policies and practices accordingly. This might involve updating guidelines for AI use in coursework, enhancing support for digital literacy among students and faculty, or developing new pedagogical strategies that leverage AI technologies in ethical and educationally beneficial ways.

Also, a key aspect of being responsive is the establishment of ongoing dialogue and feedback loops with the broader community, including those directly impacted by the integration of AI in educational contexts. Regular assessments, surveys, and forums can provide valuable insights into the effectiveness of AI applications in teaching and learning, as well as identify emerging ethical concerns and societal expectations. These insights can

then inform the iterative refinement of AI deployment strategies, ensuring they remain aligned with the evolving needs and values of the educational community.

Figure 6.2 delineates the core concepts of Responsible Research and Innovation (RRI) while illustrating how they are intricately connected. This figure serves to underscore the notion that the effective and responsible utilisation of AI tools within higher education (HE) demands a holistic approach. Such an approach integrates anticipation, reflexivity, inclusiveness, and responsiveness, fostering a framework that ensures consistent and meaningful engagement.

Anticipation allows us to foresee and prepare for both the opportunities and challenges presented by ChatGPT, setting the groundwork for a proactive stance towards digital innovation. Reflexivity encourages ongoing self-examination among all stakeholders, ensuring that the deployment of ChatGPT is continuously aligned with educational values

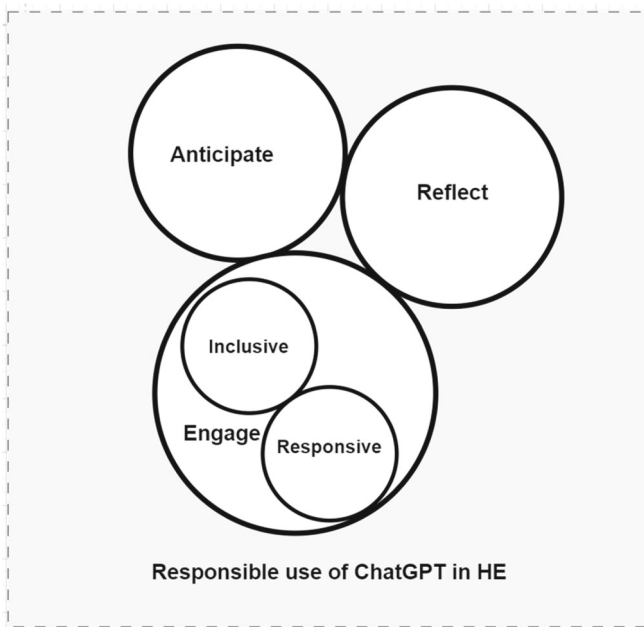


Fig. 6.2 Adapted RI framework for responsible use of ChatGPT in HE

and societal norms. Inclusiveness broadens the conversation, bringing diverse perspectives into the innovation process, thus enriching the development and application of ChatGPT in a way that respects and represents the wide array of stakeholders within the HE community. Lastly, responsiveness ensures that this approach remains dynamic and adaptable, capable of evolving with the changing needs and ethical considerations of society at large.

Together, these pillars of RRI creates a robust framework that supports the responsible integration of ChatGPT into the educational landscape. By blending anticipation, reflexivity, inclusiveness, and responsiveness, institutions can navigate the complexities of digital innovation in a manner that is both ethically grounded and socially beneficial. This holistic approach promotes not only consistent engagement with technological advancements but also ensures that such engagement is constructive, inclusive, and responsive to the diverse needs and challenges of the higher education community. An extensive RRI toolkit to guide the implementation of RRI principles across a wide range of scenarios (excluding the use of LLMs in HE) is available at <https://rri-tools.eu/>. In the next section, we propose a simple step-by-step guide that most stakeholders will find useful.

6.7 HOW TO USE GENAI RESPONSIBLY IN HE: A RESPONSIBLE INNOVATION APPROACH

When considering the use of Large Language Models such as ChatGPT for teaching and learning in higher education, it is essential to take a thoughtful approach that encompasses the principles of Responsible Research and Innovation (RRI). The process begins by **anticipating** both the positive and negative implications of such technology's deployment.

On the one hand, we must **identify the negatives**, acknowledging the risks associated with the use of GenAIs—these could include issues related to academic integrity, privacy, data security, bias, and the potential erosion of critical thinking skills. On the other hand, it is equally important to **identify the positives**, recognising the values ChatGPT brings, such as personalised learning, increased access to information, and potential for enhancing creativity and problem-solving skills.

Once we have identified these implications, we move on to **assess the controls** in place, such as existing policies and guidelines. This involves understanding at what levels these controls operate, whether at

the module, programme, or university level. If the current control mechanisms are not adequately measuring something significant, we engage in a reflective process to determine what should be measured, how, and when. This might include metrics related to usage patterns, student engagement, or learning outcomes.

The next step is to **engage with the target stakeholders**. This requires identifying who they are and how we can engage them in a manner that is inclusive and represents the diverse voices within the HE community. It is important to consider how the use of LLMs fits into larger sectoral or national frameworks that may influence HE policies and practices.

This engagement leads to **improvement**—based on feedback and reflection, we aim to enhance the implementation strategy for LLMs. This could involve updating policies, offering training for faculty and students, and developing best practices for integrating ChatGPT into the curriculum.

Finally, we circle back to **respond**. This involves looking at the metrics we've gathered and interpreting what they tell us about the impact of ChatGPT on the educational process. The metrics inform us whether the risks are being mitigated effectively and whether the values we aimed to multiply are indeed being realised. If the data indicates success, we can further optimise and expand our approach. If not, we engage in a cycle of reflection and engagement, seeking to understand how we can better align our use of ChatGPT with the principles of RRI.

In this manner, the process becomes cyclical: **anticipating** outcomes, **reflecting** on practice, **engaging** stakeholders, **responding** to evidence, and seeking continuous **improvement**. By adhering to this methodology, HE institutions can ensure the responsible use of ChatGPT, balancing innovation with ethical considerations and societal needs.

While this narrative above is hypothetical, it suggests a pragmatic and holistic approach to the responsible use of generative AI in Higher Education (HE). It can also serve as a benchmark for HE institutions to evaluate where they stand in the journey toward the responsible use of GenAI (Fig. 6.3).

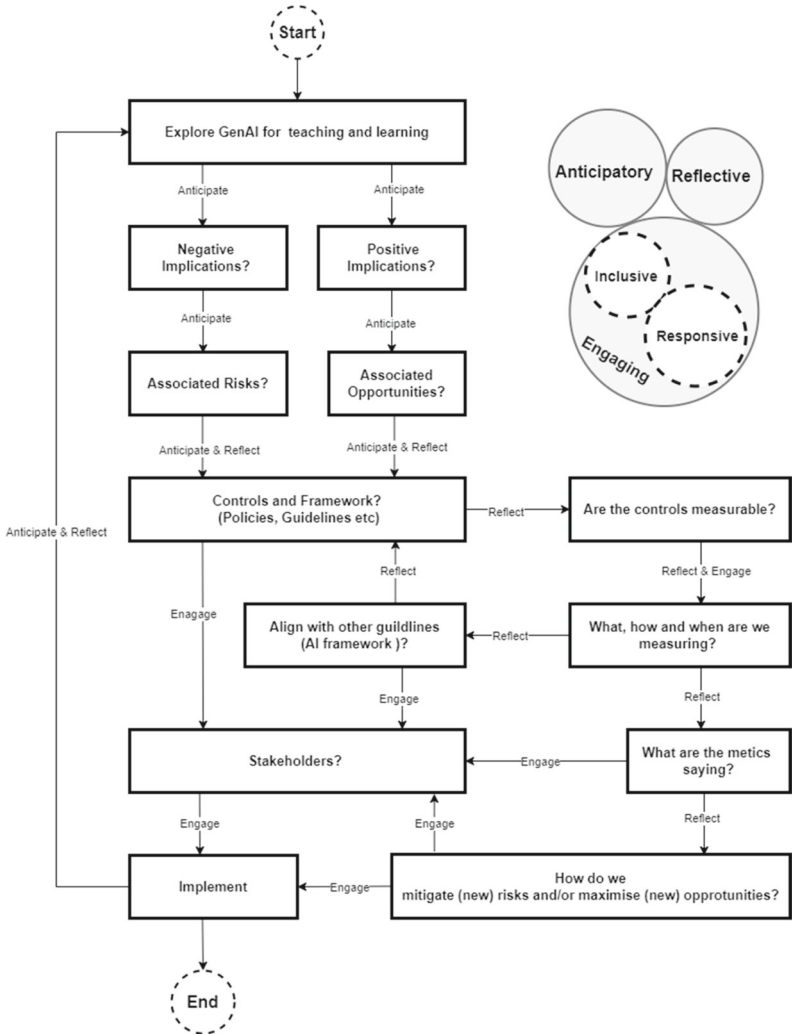


Fig. 6.3 Responsible GenAI4HE Flowchart

6.8 CONCLUSION

The integration of generative AI, like ChatGPT, into higher education (HE) presents an unparalleled opportunity to enhance the teaching and learning ecosystem. This chapter has traversed the multifaceted landscape of utilising generative AI within HEIs, emphasising the importance of responsible innovation to navigate the ethical, social, and technical complexities inherent in this digital transformation.

The unfolding interest among educators to incorporate tools like ChatGPT into their pedagogical practices, as revealed through surveys and discussions, underscores a collective aspiration to leverage AI's potential. However, this enthusiasm is tempered by legitimate concerns surrounding academic integrity, equity, and the preservation of critical thinking skills. The disparity between the interest in and actual usage of generative AI technologies in education highlights a cautious approach, prioritising responsibility over rapid adoption (Stahl, 2021; Stahl et al., 2023).

In addressing these concerns, the chapter has outlined a Responsible Research and Innovation (RRI) framework that offers a holistic approach to the ethical integration of generative AI in HE. This framework comprises four cornerstone principles: anticipation, reflexivity, inclusiveness, and responsiveness, each playing a pivotal role in fostering an ethical, equitable, and effective educational environment.

- **Anticipation** calls for a proactive stance, preparing for both the challenges and opportunities presented by generative AI. This forward-looking perspective is not about predicting the future with precision but about being prepared for various plausible futures. The anticipation of risks such as bias, misinformation, and academic integrity issues necessitates a multifaceted approach to mitigation, emphasising the need for continuous dialogue and adaptation.
- **Reflexivity** demands a critical self-examination of the roles, assumptions, and actions of all stakeholders involved in the innovation ecosystem. By reflecting on the broader context in which AI is deployed, educators, policymakers, and technologists can align their practices with societal values and educational goals. This continuous process of reflection ensures that the use of generative AI remains aligned with the principles of responsible innovation.

- **Inclusiveness** stresses the importance of engaging a diverse range of stakeholders in the innovation process. By democratising the conversation around AI in education, we can ensure that a broad spectrum of perspectives and values is considered, enriching the development and application of these technologies. Such an inclusive approach not only addresses concerns of equity and accessibility but also leverages the collective intelligence of the community to steer innovation towards societal benefit.
- **Responsiveness** highlights the need for innovation systems to be adaptable, and capable of responding to emerging societal needs and ethical considerations. This requires mechanisms for integrating societal values into innovation practices, ensuring that generative AI tools like ChatGPT are used in ways that reflect educational goals and societal expectations.

The intersection of these principles forms the foundation for a pragmatic Responsible Innovation approach to using GenAIs in HE. By embracing these guiding principles, HEIs can navigate the complexities of digital innovation, ensuring that the deployment of generative AI technologies like ChatGPT is both responsible and beneficial.

As we look to the future, the path forward requires not only adherence to these principles but also an ongoing commitment to dialogue, collaboration, and adaptation. The dynamic nature of technology and society means that what constitutes responsible use today may evolve tomorrow. Therefore, HEIs must remain vigilant, continuously assessing and adjusting their practices in response to new insights, challenges, and opportunities.

In conclusion, the responsible use of generative AI in higher education is not merely an aspirational goal but a necessary commitment to ensuring that technological innovation serves to enhance, rather than diminish, the quality and integrity of education. By fostering an environment of anticipation, reflexivity, inclusiveness, and responsiveness, we can harness the transformative potential of generative AI to create an educational landscape that is equitable, ethical, and aligned with the enduring values of academic excellence. The journey towards responsible innovation is ongoing, and each step forward enriches our collective understanding and capability to shape a future where technology and humanity converge in the service of education and societal progress.

REFERENCES

- Margherita, N., & Bernd, S. (2018). RRI in higher education. *The ORBIT Journal*, 1(4), 1–8. <https://doi.org/10.29297/orbit.v1i4.78>
- OpenAI Community. (2023). *Exploring emotional connections in human-AI interactions*. Retrieved from <https://community.openai.com/t/exploring-emotional-connections-in-human-ai-interactions/281770>
- Stahl, B. C. (2021). *Artificial intelligence for a better future an ecosystem perspective on the ethics of AI and emerging digital technologies*. <http://www.springer.com/series/13811>
- Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). *Ethics of artificial intelligence case studies and options for addressing ethical challenges SpringerBriefs in research and innovation governance* (Vol. 1).
- Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, 42(9), 1568–1580. <https://doi.org/10.1016/j.respol.2013.05.008>
- Tarisayi Kudzayi. (2024). ChatGPT use in universities in South Africa through a socio-technical lens. *Cogent Education*, 11(1), 2295654. <https://doi.org/10.1080/2331186X.2023.2295654>
- Xiao, P., Chen, Y., & Bao, W. (2023). *Waiting, banning, and embracing: An empirical analysis of adapting policies for generative AI in higher education*. <https://www.celt.iastate.edu/resources/ai-teach-learn/>