

WHITE PAPER

Community of Practice on AI in Higher Education

How to Train Your GPT

Ensuring Your Voice in Your Research

Session 3 | 20 August 2026

Convened by Dr Susheila Moodley (PhD)

University of Johannesburg

College of Business and Economics

Guest Presenter: Professor Chatradari Devroop, Research Professor, Faculty of Arts and Design

Tshwane University of Technology

Sponsored by the College of Business and Economics, University of Johannesburg

Professor Thea Tselepis, PhD, Associate Professor and Vice Dean: Teaching and Learning, College of Business and Economics, University of Johannesburg. Co-host and Sponsor

1. Executive Summary

On 20 August 2026, the Community of Practice (CoP) on AI in Higher Education, convened by Dr Susheila Moodley (PhD), Academic Director, FutureBanker, and Professor of Practice, College of Business and Economics, University of Johannesburg, held its third session of the year under the title "How to Train Your GPT: Ensuring Your Voice in Your Research." The session brought together academics, postgraduate supervisors, and researchers from across South African higher education institutions to address a question that is increasingly urgent: how do researchers use generative AI responsibly in the research process while preserving their epistemic voice, disciplinary integrity, and accountability?

This white paper synthesises the preparatory literature review conducted across five empirical studies from China, South Africa, and Pakistan (2024 and 2025), the cross-document thematic analysis that informed the session, the live presentation and workshop delivered by Professor Chatradari Devroop of Tshwane University of Technology, and the discussion that emerged from participants during the session.

The paper advances three core arguments. First, the psychological experience of using AI in research, characterised by dissonance, hesitation, and the need for coping strategies, is cross-contextually consistent and represents a structural feature of AI's arrival in high-stakes academic work, not a culturally specific reaction. Second, the institutional response that reduces harm is also cross-contextually consistent: clear policy and skills training, not detection and prohibition. Third, what is specifically absent in the African and South African case is the empirical research base itself: not the phenomenon, but the localised evidence of how it plays out inside African institutional, infrastructural, and epistemic conditions.

2. Purpose and Scope

This white paper serves as the formal output of the third CoP session of 2026, a joint initiative of FutureBanker and the University of Johannesburg, College of Business and Economics. Its purpose is threefold: to document the evidence base reviewed in preparation for the session; to capture the key arguments, frameworks, and practical recommendations that emerged from the session itself; and to provide a resource that CoP members and the broader South African higher education sector can use to inform institutional policy, supervisory practice, and postgraduate training on the ethical and responsible use of AI in the research process.

The scope is deliberately focused on AI in the research process, as distinct from AI in teaching, learning, or assessment, which were the subjects of earlier CoP sessions and the prior white paper on AI grading. The questions driving this session were: What does the literature tell us about how researchers are actually using generative AI? What psychological and ethical tensions arise? What institutional conditions make responsible use more or less likely? And how do researchers preserve their voice when working with tools trained predominantly on Global North knowledge systems?

3. The Literature: What Scholars Are Saying About AI in Research

In preparation for the session, five empirical studies published in 2024 and 2025 were reviewed and summarised by Dr Moodley. These studies spanned three national contexts (China, South Africa, and Pakistan) and shared a common feature: in each case, the researchers and students studied were working in English as a second language within evaluation systems that carry high stakes. The thematic analysis across these sources yielded five convergent themes.

3.1 Theme 1: The Dissonance Is Real and Structural

Across every context studied, researchers and students describe a genuine internal conflict when using AI in research. Ren, Zheng and Zhang (2025), studying 28 master's students at three Chinese universities using grounded theory, identified cognitive dissonance as the core category unifying their findings: strong intention to use ChatGPT, driven by subjective norms and technological expectations, collides with barriers to effective use, producing psychological discomfort. Rafi and Amjad (2025), working with 86 doctoral scholars and 7 supervisors in Pakistan, report the same tension in different language: hope for efficiency and creativity alongside fear of hallucination, ethical exposure, and loss of critical thinking. Muringa's South African students describe caution and self-monitoring rather than confident adoption.

The shared finding: nobody is simply "using AI." Everyone is negotiating a relationship with it under conditions of uncertainty. This appears to be a structural feature of researchers encountering generative AI inside high-stakes academic evaluation systems, wherever it has been studied.

3.2 Theme 2: Institutional Ambiguity Amplifies Harm

The single most consistent finding across the four primary empirical studies is that the vagueness of institutional response to AI, not AI itself, produces the ethical and psychological strain. Chinese universities issued ambiguous notices that theses "would be screened by AIGC-detection tools" without defining acceptable use, heightening anxiety more than either clear permission or prohibition. South African students, absent formal guidelines, set their own individual boundaries for AI use, producing wide inconsistency across the student body and leaving them underprepared and reliant on informal networks. Pakistani doctoral scholars called explicitly for institutional AI literacy training and clear policy, naming the absence of both as a source of risk.

The literature supports assessment redesign and clear policy, not detection and prohibition. This corroborates the position established in the CoP's earlier white paper on AI grading.

3.3 Theme 3: AI Is Filling a Supervision Gap

Both the Modiba, Van den Berg and Mago (2025) review and the Rafi and Amjad (2025) study independently describe AI being used as a substitute for supervisory capacity, not merely a supplement to it. Modiba et al. state that "AIs provide a supporting environment during the research journey especially where supervisors do not have capacity to scaffold research candidates." Rafi and

Amjad's participants describe AI "acting as a digital peer or supervisor, especially when human supervisors are unavailable."

This finding requires careful contextualisation. In under-resourced Global South contexts, AI-as-supervisor-substitute may reflect and entrench pre-existing supervision capacity shortages rather than solve them. This is materially different from a well-resourced Global North university where AI supplements an already-available supervisor. The literature does not yet distinguish these cases, and this distinction is critical for South African higher education policy.

3.4 Theme 4: Context, Language, and Epistemic Mismatch

Rafi and Amjad's Pakistani scholars report that generative AI "literalizes idioms and mostly misses their cultural connotations" when working with local material. Muringa's South African students navigate AI tools built and trained largely outside their epistemic and linguistic context. Modiba et al.'s review cites Kiemde and Kora's (2022) argument that AI ethical frameworks should not rely solely on a Western vision and should instead incorporate African values.

Generative AI tools carry an implicit epistemic default that is not neutral. Researchers working outside that default experience friction that researchers working inside it do not. This is not simply a language-translation problem; it is a deeper issue of whose knowledge patterns the tool was trained to recognise as valid.

3.5 Theme 5: Coping Strategies Converge Across Contexts

Despite very different national and institutional settings, the practical strategies researchers adopt to manage AI responsibly are strikingly consistent: improving prompt quality and prompt literacy; feeding the model relevant domain-specific material rather than relying on its base knowledge; verifying AI output against primary or academic sources; maintaining supervisor involvement as the final check on critical engagement; and developing the capacity to recognise AI-generated text by its stylistic signatures.

This convergence is significant: "training your AI" is not a Global North practice being imported. It is a practice independently arrived at by researchers in China, Pakistan, and South Africa as the functional response to the same underlying problem.

3.6 Theme 6: The African Empirical Base Remains Thin

Modiba, Van den Berg and Mago's (2025) systematic review is the starkest evidence. From 62 initially identified articles, only two ultimately addressed generative AI supporting research in African higher education. The authors state directly that sourcing scientific articles was difficult because "search engines could not find sufficient literature on this topic." This aligns with Bond et al.'s (2024) meta-systematic review finding that voices from the Global South are systematically under-represented in the AI-in-higher-education literature.

The absence of African-authored empirical research on AI in the research process is itself a significant finding, not a limitation of the search strategy. It is the argument, not the gap.

4. The Session: Key Arguments from Professor Devroop

Professor Chatradari Devroop, Research Professor in the Faculty of Arts and Design, Tshwane University of Technology, delivered a structured workshop on the ethical and responsible use of AI in the research process. His presentation was grounded in practical experience of integrating AI across postgraduate supervision, faculty development, and institutional policy at a South African university of technology. Prof. Devroop opened with a historical preamble, tracing the evolution of knowledge dissemination from manuscript culture through Gutenberg's printing press to the digital era, and noting that AI has been in development for over seventy years, dating to the Dartmouth Summer Research Project of 1956 and long predating the public launch of ChatGPT in November 2022. He also surveyed institutional responses to AI, identifying three categories: full embrace (as at Arizona State University and MIT), cautious adoption with policy development, and outright prohibition. The key arguments and frameworks presented are summarised below.

4.1 AI as Agent, Not Tool

Prof. Devroop then established a foundational conceptual distinction: AI should not be understood as a tool but as an agent. A tool (a hammer, a pen) cannot act on its own. Generative AI exercises a delegated form of agency: once instructed, it can generate, compose, and create outputs autonomously, without the researcher specifying each step. The claim is not that AI possesses intention or understanding, but that the researcher delegates part of the productive act while remaining answerable for what returns. This distinction matters because it reframes the researcher's responsibility. Working with an agent requires ongoing oversight, verification, and accountability in a way that working with a passive tool does not.

4.2 The Five Core Habits of Accountable Research

Prof. Devroop proposed five verbs as a framework for responsible AI use in research, offered as a mantra for researchers and postgraduate students:

Decide: when it is appropriate to use AI and when it is not.

Protect: sensitive data, ensuring it remains within closed institutional systems and is not uploaded to public platforms.

Prompt: with accountability, using bounded, context-specific instructions rather than open-ended requests.

Verify: every output, treating all AI-generated content as a first draft requiring triangulation against independent sources.

Disclose: transparently, specifying what tool was used, for what task, what data was fed in, how outputs were verified, and confirming that the researcher retains full responsibility for the results.

4.3 The Seven Responsible Principles

Complementing the five core habits, Prof. Devroop presented seven responsible nodes that researchers must observe when working with AI. He described these principles as "traffic rules" that "are not there to simply slow us down. They are simply there to keep us from crashing with the kind of work that we are doing."

The seven nodes are: human-centredness (ensuring humans remain at the centre of the research process); safety (providing safety for oneself, one's institution, and the research being undertaken); reliability (verifying the reliability of AI-generated information, recognising that AI is "only a pattern recognition system"); transparency (declaring what was done, how it was done, and how information was obtained); accountability (being able to defend one's work, with viva voce examinations becoming increasingly important as a mechanism for ensuring this); fairness (treating data, sources, and research participants equitably, and guarding against the reproduction of bias in AI outputs); and privacy (protecting personal and institutional information throughout the research process).

The seven responsible nodes and the five core habits are presented as complementary frameworks: the habits address what the researcher does with AI, while the nodes address the principles that govern how the researcher approaches the work.

4.4 Fluency Is Not Truth

A recurring emphasis throughout the session was the distinction between fluency and truth. AI generates probable output based on learned patterns; it does not understand, verify, or contextualise. Its outputs are confident and linguistically polished, which creates the risk that users confuse fluency with accuracy. Prof. Devroop illustrated this with the concept of hallucination, that is, AI's tendency to fabricate citations, statistics, or claims when it cannot locate appropriate pattern completions, and emphasised that every AI output must be treated as a first draft requiring human verification.

Prof. Devroop described AI as "an autocomplete on steroids". The phrase captures both the speed of the output and its shallowness: the system predicts what should plausibly come next without understanding or knowing anything. He cautioned that AI does not observe disciplinary boundaries, drawing language and concepts from across disciplines without contextual awareness. He also identified what he called the "Amplification Effect": AI's tendency to amplify and embellish, always presenting a grand version of its output and defaulting to positive affirmation rather than critical assessment. These characteristics make human verification not merely advisable but essential.

4.5 The Western Epistemic Default

Prof. Devroop located AI's epistemic bias historically. The bulk of material available on the internet, the corpus on which large language models are trained, reflects Western knowledge systems, because these were the systems that moved from print to digital earliest and most extensively. African, Asian, and indigenous knowledge systems, which were predominantly oral and passed down through generations rather than printed and digitised, are under-represented in training data. The practical consequence: when a South African researcher asks AI to generate content, the output defaults to

Western epistemological frames unless the researcher actively intervenes by providing context and domain-specific sources.

Prof. Devroop situated this argument within South Africa's specific historical context, noting the country's experience of double colonisation (Dutch from 1652, British from 1806) and cautioning against treating Africa as a monolith. He outlined the key South African frameworks that researchers must observe: the Protection of Personal Information Act (POPIA), individual university policies, the Universities South Africa (USAf) guidelines, and the Academy of Science of South Africa (ASSAf) guidelines for epistemic justice and African scholarship representation.

This argument converged directly with the literature review finding on epistemic mismatch and reinforced the CoP's position that localised large language models, or at minimum localised prompting practices, are necessary for responsible AI use in African higher education research.

4.6 Practical Prompting: The Bounded Prompt Formula

Prof. Devroop provided participants with a practical prompting formula designed to preserve researcher voice and ensure accountability:

State the task and context: define the research domain, the institutional setting, and the disciplinary frame before issuing any instruction to the AI.

Set boundaries: explicitly instruct the AI not to invent citations, not to fabricate data, and to work within the sources provided.

Feed domain-specific material: upload relevant articles, datasets, or field-specific literature into a closed project space so the AI works from verified sources before defaulting to its general training data.

Define the output format: request bullet points, outlines, or structured summaries rather than finished essays, so the researcher's own writing and argumentation remain the primary intellectual contribution.

Verify through triangulation: cross-check AI outputs across multiple platforms, including at least one non-Western model, and against primary academic sources.

Prof. Devroop recommended triangulating AI outputs across multiple platforms, describing his own practice of using one Western model (Perplexity or Claude) and one Chinese model (Manus or DeepSeek) to surface epistemic differences. He noted that Perplexity is comparatively reliable for source verification because it surfaces its references: "Perplexity is the one platform that never makes up stuff. It will tell you exactly, I got this from this article on this page on this place." He connected this to the research concept of triangulation: using at least two, and preferably three, separate AI platforms and checking whether they converge on the same answer. The claim in the quotation above should be read as a relative judgement about citation transparency rather than a guarantee of accuracy, since retrieval-based systems still misattribute and mis-summarise. The verification principle set out in 4.2 applies to them in full.

4.7 POPIA and Research Data

Prof. Devroop addressed the Protection of Personal Information Act (POPIA) directly, establishing clear boundaries for what must never be entered into public AI platforms: identifiable participant information, sensitive personal data, unpublished institutional datasets, and any data that could identify individuals or compromise institutional confidentiality. He noted that closed-system options existed at the time of the session within platforms such as Microsoft Copilot (work or enterprise mode) and Claude (project spaces, with data-retention and model-training controls), but that researchers must actively configure these settings rather than assuming default privacy protections, and must confirm the terms attached to their own institutional licence.

Prof. Devroop clarified that paid versions of AI platforms are subject to the same POPIA requirements as free versions: "The paid version simply allows you faster access, larger data platforms, wider usability and larger sized documents. Irrespective, you are not allowed to upload [personal information]." However, researchers can use AI for processing and analysis of institutional data provided they work within a closed system to ensure data is not made publicly available or used for model training.

4.8 When Not to Use AI

The session established that responsible practice sometimes means not using AI at all. Specifically: for processing confidential data; for high-stakes decisions including judicial or ethical judgements; for any task where the researcher cannot independently verify the evidence; and for outsourcing core intellectual analysis. The analysis must be done by the human researcher.

Prof. Devroop was emphatic about AI's limitations: AI does not understand ethics and cannot take responsibility for research outcomes. He noted that AI cannot replace the critical guidance of a supervisor, who brings deep disciplinary knowledge, and cautioned against students relying on AI for book summaries, which produce "smoothed" versions that strip out important contextual details. Where a university has explicitly restricted AI use, that restriction must be observed.

5. Discussion: Themes from the Room

Participant engagement during the session surfaced several themes that extend and complicate the findings from the literature review.

5.1 The Question of Agency and Truth

A participant raised a philosophically significant challenge: if AI is being continuously fed knowledge, then the concern about AI having no guarantee of truth is itself grounded in the paradoxes and contradictions of human knowledge. The question sharpened when paired with the concept of agency: when AI begins to decide what is true and what is false, the implications become existential rather than merely practical. Prof. Devroop's response, that truth is always contested and that researchers must specify their epistemological frame to the AI rather than assuming it shares one, is itself a pedagogical principle worth embedding in postgraduate training.

5.2 Sensitive Research and Ethical Courage

A substantial discussion emerged around whether the emphasis on ethics and responsible AI use risks discouraging researchers from pursuing sensitive or politically charged topics. The consensus was clear: no research topic is off limits, but every research topic requires ethical rigour. The distinction is between the question (which should be fearless) and the method (which must be accountable). Anonymisation, contextualisation, and returning findings to communities were identified as non-negotiable practices regardless of whether AI is involved. When research findings involve comparisons across racial or demographic groups, Prof. Devroop recommended presenting data using neutral labels (racial group A, B, C, D) rather than identifying groups by name, allowing the research to inform interventions while protecting the dignity and privacy of the communities studied.

5.3 The Positive Bias of the Room

When Prof. Devroop asked participants to offer one word they associate with AI in research, every response was positive: efficiency, assistant, productivity, brainstorming, paraphrase, time saver. No participant offered a negative association. This is itself a finding: the room's practitioners have moved past the fear stage and into active, if cautious, adoption. The challenge now is not persuading academics to use AI but equipping them to use it well.

5.4 Model Improvement and Infrastructure

A discussion initiated by Dr Moodley addressed whether the continuous improvement of AI models reduces the risk of model collapse. Prof. Devroop confirmed that AI models are being updated continuously, with improvement trajectories that have so far been steep, and offered a personal projection that the rate of improvement would begin to taper around 2040. On the infrastructure side, he noted that data centres consume enormous quantities of water and energy for cooling, and that siting decisions are driven principally by the availability of cheap power, land, and connectivity, with several large facilities now concentrated in the Gulf states on that basis. Water scarcity operates as a constraint on such projects rather than a reason for them, which sharpens the question for South Africa: any move to build local data centres in support of localised models would have to be weighed against existing water and electricity pressures.

He also observed that AI is learning from the indigenous knowledge being introduced by users, and that models will continue to grow as more countries and cultures engage with them. He connected this to the ethical dimension: "This is more around ethics now. And this is the decision that politicians need to make. We need to start to develop a sense of control to rein AI in, to make sure that it is always human-centred." This forward-looking discussion reinforced the white paper's argument that institutional and governmental engagement with AI governance is not optional but urgent.

6. Recommendations

Drawing on the literature, the session, and the discussion, the CoP advances the following recommendations for South African higher education institutions, supervisors, and researchers.

6.1 For Institutions

Develop explicit, context-sensitive AI use policies that define acceptable use rather than defaulting to detection and prohibition. The literature across three national contexts confirms that ambiguous institutional signals amplify researcher anxiety and produce inconsistent practice.

Invest in AI literacy training for both postgraduate students and supervisors. The training should cover prompt literacy, verification practices, POPIA compliance in AI-mediated research, and the epistemic limitations of large language models trained on predominantly Western corpora.

Require transparency in AI use rather than simple disclosure. A disclosure statement should specify what tool was used, for what task, what data was provided, how outputs were verified, and that the researcher retains full responsibility for the results.

Ensure that closed-system AI access is available to researchers handling sensitive or confidential data, through institutional licensing of platforms that offer work-mode or project-space configurations with training-data opt-out.

6.2 For Supervisors

Recognise that AI is already filling supervisory gaps, and respond by increasing supervisory engagement rather than assuming AI reduces the need for it. The literature shows that students use AI as a substitute when supervisors are unavailable; the institutional response should be to address availability, not to ban the substitute.

Integrate AI literacy into the supervision relationship. Supervisors should be equipped to guide students in prompt construction, output verification, and the recognition of AI-generated text, and to model these practices in their own research.

Maintain the viva voce or oral defence as a critical assessment mechanism. When a student must verbally defend their work without notes, the depth of their engagement with their own research becomes visible in ways that written submissions alone cannot reveal. Prof. Devroop emphasised that viva voce examinations are becoming increasingly important as a mechanism for ensuring accountability: when a student must explain their process orally, the authenticity of their engagement with the research, including their use of AI, becomes transparent.

6.3 For Researchers

Adopt the five-habit framework (Decide, Protect, Prompt, Verify, Disclose) as a minimum standard for responsible AI use in research.

Treat every AI output as a first draft. Verify all dates, statistics, citations, and claims against independent sources. Triangulate across platforms, including at least one non-Western model, to surface epistemic bias.

Prof. Devroop emphasised that simple disclosure of AI use is insufficient. The five-question disclosure template set out in 4.2 should therefore appear within the methodology section itself, where it can be examined as part of the research design, rather than being relegated to an acknowledgements note.

Feed the model your context before asking it to generate. Provide your disciplinary frame, institutional setting, theoretical positioning, and domain-specific sources. Without this, the AI defaults to a Western epistemic frame that may not serve your research.

Maintain an AI use log documenting how AI was used at each stage of the research process. This log becomes the basis for a transparent and accountable methodology section.

Remember that responsible practice sometimes means not using AI at all, particularly for confidential data, core intellectual analysis, and any task where verification is not possible.

7. Closing Reflection

The title of this session, How to Train Your GPT: Ensuring Your Voice in Your Research, contains an assumption worth making explicit. Training the AI is not a one-time configuration. It is an ongoing practice of feeding context, verifying output, and exercising judgment. The researcher's voice is not something that survives AI use passively; it is something that must be actively constructed and defended through the way the researcher engages with the tool.

The literature reviewed for this session, and the discussion that emerged from it, converge on a single insight: the dissonance researchers experience when using AI is not a problem to be eliminated but a signal to be attended to. It indicates that the researcher is still thinking critically about the relationship between their own knowledge and the AI's output. When that dissonance disappears, when the researcher stops questioning, stops verifying, stops contextualising, that is when voice is lost.

The work of this CoP, and of the institutions it serves, is to ensure that the dissonance remains productive: that it drives better practice rather than paralysis, and that the institutional conditions surrounding AI use (clear policy, skills training, supervisory engagement, and equitable access) make it possible for every researcher to use this technology responsibly, ethically, and in service of knowledge that reflects their own intellectual contribution.

The intellectual gap is real. The work to close it has to be ours. – Dr Susheila Moodley, LinkedIn, 2025

References

- Bond, M., Khosravi, H., De Laat, M., et al. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education*, 21, Article 38. <https://doi.org/10.1186/s41239-023-00436-z>
- Kiemde, S. M. A., & Kora, A. D. (2022). Towards an ethics of AI in Africa: Rule of education. *AI and Ethics*, 2(1), 35–40. <https://doi.org/10.1007/s43681-021-00106-8>
- Modiba, F. S., Van den Berg, A., & Mago, S. (2025). Opportunities and challenges of generative artificial intelligence supporting research in African classrooms. *South African Journal of Higher Education*, 39(3), 173–193. <https://dx.doi.org/10.20853/39-3-6272>
- Muringa, T. P. (2025). Exploring ethical dilemmas and institutional challenges in AI adoption: A study of South African universities. *Frontiers in Education*, 10, 1628019. <https://doi.org/10.3389/feduc.2025.1628019>
- Rafi, M. S., & Amjad, I. (2025). The role of generative AI in writing doctoral dissertation: Perceived opportunities, challenges, and facilitating strategies to promote human agency. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-025-13476-x>
- Ren, X., Zheng, W., & Zhang, M. (2025). Navigating cognitive dissonance: Master's students' experiences with ChatGPT in dissertation writing. *Frontiers in Psychology*, 16, 1542559. <https://doi.org/10.3389/fpsyg.2025.1542559>

Community of Practice on AI in Higher Education

University of Johannesburg | aiiineducation.co.za

FUTUREBANKER

© 2026. All rights reserved.