The Hague Centre
for Strategic Studies

Think First, Prompt Second

HCSS Code of Conduct for the Responsible Use of Generative AI

Jesse Kommandeur, Sofia Romansky, Rens van Dam, Thijs van Aken and Tim Sweijs

August 2026

The Hague Centre
for Strategic Studies

Think First, Prompt Second

HCSS Code of Conduct for the Responsible Use of Generative AI

August 2026

Authors: Jesse Kommandeur, Sofia Romansky, Rens van Dam, Thijs van Aken and Tim Sweijs.
Contributors: the HCSS research team

First Version: October 2024
Current Version: August 2026

Cover: Canva

This publication is **AI-supported**. The authors take full responsibility for the integrity of the work.

The observations presented in this manual are the result of in-house independent research and interviews. Responsibility for the content rests with the authors and the authors alone.

Table of Contents

1.	**Introduction**	**1**
2.	**Guidelines**	**4**
2.1.	Cultivate Scepticism	4
2.2.	Practice Active Learning	5
2.3.	Consider Fit Between Tool and Purpose	6
2.4.	Safeguard Integrity	7
2.5.	Maintain Confidentiality	8
2.6.	Use Deliberate Prompting	10
2.7.	Ensure Transparency	10
2.8.	Accommodate Experimentation	12

Key Takeaways

Cultivate Scepticism
Researchers should treat generative AI outputs as provisional and critically evaluate them in light of known limitations such as bias, hallucination, and lack of explainability.

Practice Active Learning
Generative AI should be used to support reflective and iterative learning processes, not to replace critical thinking or independent human assessment.

Consider Fit Between Tool and Purpose
Different generative AI tools have distinct strengths, and selecting the appropriate tool for each task is essential for producing reliable and relevant outcomes.

Safeguard Integrity
The use of generative AI must be deliberate, with researchers retaining full responsibility for verifying outputs and ensuring methodological rigour.

Maintain Confidentiality
Sensitive, personal, or confidential information must not be shared with generative AI unless secure and compliant environments are guaranteed.

Use Deliberate Prompting
Effective use of generative AI depends on clear intent, structured prompting, and continuous refinement of inputs to guide reliable outputs.

Ensure Transparency
The role of generative AI in research workflows should be documented and communicated within teams and, where relevant, in outputs.

Accommodate Experimentation
Researchers should actively experiment with generative AI tools and share best practices to build institutional knowledge and capability.

1. Introduction

Artificial intelligence (AI) has become omnipresent in everyday life. With recent and ongoing advances, generative AI has become a technology that can offer practical support across many domains by processing vast amounts of data and generating outputs ranging from text to images, audio and code. For knowledge institutes, these capabilities open a variety of methodological possibilities while raising new professional responsibilities. This code of conduct focuses on the role that specifically generative and agentic AI can play as a methodological tool within knowledge institutes. How knowledge institutes should engage with these tools, however, remains contested. Critics of AI adoption point to its tendency to hallucinate information, the biases inherited from training data, and the deterioration of human skills that interaction with AI can bring.

AI-enthusiasts, by contrast, emphasise the technology's potential to accelerate research and innovation, surface patterns in complex data, and expand analytical capabilities beyond what individuals can achieve alone. As a knowledge institution, HCSS seeks to strike a balance between these two perspectives in order to both reap the benefits and mitigate the downsides.

The current version of the **HCSS Code of Conduct for the Responsible Use of Generative AI** builds on previous iterations of the Code of Conduct including the last one published in October 2024, a review of in-house (good and bad) practices, multiple dedicated discussions with the HCSS research staff, insights from a workshop hosted with researchers and policymakers in November 2025, and an expert roundtable in April 2026. The roundtable brought together computer scientists, academic researchers, philosophers of science, geopolitical analysts, and practitioners working on AI policy. Across three sessions, participants discussed the technical frontier of generative AI, its implications for research practice, and its geopolitical dimensions. The conclusions of these discussions have been incorporated throughout the Code of Conduct.[1]

For knowledge institutes, which aim to inform not only the broader public but also public and private policy making, the integrity of information and validity of claims is non-negotiable. In the age of generative AI, where any individual regardless of educational background has access to near-limitless pre-digested information, a key role remains for knowledge institutes. They must preserve human critical thinking and analysis, as well as the development of these capabilities. Generative AI can be used as an instrument to ensure that researchers maximise their use of existing knowledge. Additionally, AI-based methods can collect, classify and synthesise large bodies of unique information, making unstructured data digestible for human interpretation and analysis. As such, it can enable knowledge institutes to focus human effort on higher-value work, producing original insights rather than restating what already exists and innovating how knowledge is shared and taught.

[1] Given the pace at which generative AI is evolving, both technologically and in its societal embedding, HCSS intends to convene this roundtable on an annual basis. The Code of Conduct will be reviewed and updated in light of each subsequent session, so that it continues to reflect both the state of the technology and the evolving practice of responsible use within knowledge institutes.

Table 1 below presents different techniques through which the use of generative AI can contribute to workflows in different research phases, including brainstorming, research and system design, data collection, analysis, reporting and visualisation.

Table 1. Use of AI throughout research phases

	Creating knowledge products Writing & analysis	Creating knowledge systems Data Science & Coding
Research phase/ technique	**Prompt engineering techniques**	**Agent engineering techniques**
Brainstorming	• Open-ended instructions • Zero-shot prompting • Meta-prompting	• Agent-assisted collaborative scoping • Delegated feasibility assessment • Socratic prompting
Research Design / System Design	• Task-specific instructions • Few-shot prompting (indicate examples of what you expect) • Context and function of task	• Instruct agent with brainstorm results • Elicitation prompting • Iterative plan improvement
Data Collection	• Sourcing and explainability • Generate knowledge prompting • Retrieval augmented generation	• Agent-assisted data exploration • Agent-annotator on sample data • Human-In-The-Loop (HITL) quality assurance • Retrieval augmented generation
Analysis	• Chain of thought prompting (outline specific steps) • Self-consistency and prompt chaining • Active-prompting	• LLM-as-a-judge • Collaborative benchmarking • HITL quality assurance
Reporting	• Reflexion • Self-consistency and prompt chaining • Active-prompting • Data visualisation • Graphic design	• Agentic Reflexion • Agent-assisted technical documentation • AI content and process labelling • Tailored content generation: text, audio and video

While generative AI is not something to avoid, the efficiency it offers is not a substitute for effectiveness. For the productive and pragmatic integration of AI into knowledge-generation workflows, individual and institutional behavioural as well as technical best practices are needed. Familiarity with AI techniques, as well as the adoption of a mindset cognisant of a specific tool's strengths and limitations, form an indispensable part of research literacy to make generative AI a supplement that supports discreet research stages, rather than a shortcut to finished outputs. Strong research still depends on precisely formulated questions, rigorous research designs, and critical evaluation of claims based on empirical evidence. Overreliance on automated outputs can lead to poor outputs while also creating unhelpful habits that hamper learning and development: superficial engagement with content, reduced methodological clarity, or premature confidence in unverified summaries.

Generative AI should therefore not replace critical thinking. Rather, it should enable researchers (including both research analysts and data scientists) to augment their work. Research projects that were impossible to execute only a few years ago, have now become feasible because of the analytical capabilities of generative AI. This capability extension requires an additional effort though: it requires us to cultivate scepticism, actively learn new methods and workflows, and foster new forms of research integrity. Critical thinking remains crucial along all research stages whilst working in 'autopilot' mode should be avoided at all times. Continuous experimentation and learning from best or good practices is necessary

to keep pace with the technology as it develops. All of this must further be facilitated within a secure technical architecture for generative AI use as well as an AI research literacy amongst analysts.

The guidelines in this Code of Conduct elaborate the responsible use of generative AI within HCSS. These recommendations are oriented primarily towards LLMs, which are the dominant form of generative AI in research workflows at the time of writing. The frontier of the field is, however, moving towards categories of models, including so-called world models, that are likely to raise questions distinct from those addressed here. Awareness of these developments and preparedness to revisit the recommendations in this document in light of them, is part of the ongoing responsibility that this Code of Conduct describes, and one of the reasons why HCSS will revisit and update this document annually.

2. Guidelines

Responsible use of generative AI necessitates both responsible technical solutions and responsible behaviour. Researchers must not only learn the best methods for interacting with AI-models but also adopt a particular mindset and concomitant practices. Crucially, these changes cannot be effectuated without broader organisational support; workflows and culture need to be such that they encourage researchers to engage with generative AI in responsible ways, rather than inadvertently create incentives for ultimately harmful behaviour. At the same time, it is important to acknowledge that very limited adoption of generative AI into workflows may not grant the desired benefits. This section presents eight recommendations for researchers to be able to optimally make use of generative AI within their workflows. These recommendations provide a cohesive picture of the ways in which researchers can, and ought to, interact with generative AI to further the goals of knowledge institutes and continue to contribute to the creation and dissemination of epistemologically sound knowledge. For each recommendation, we provide some practical reflection on what it would mean for daily work at HCSS, which may also extend to other research institutes.

2.1. Cultivate Scepticism

Education is the fundamental building block of effective interactions between human researchers and generative AI. Due to the inherent architecture of generative AI models, all readily accessible tools have certain technical limitations, as presented in Table 2 below.[2]

Table 2. Limitations of generative AI models

Limitation	Description	Cause
Errors and sycophancy	Generative AI models can generate plausible sounding but factually incorrect or fabricated outputs, as well as sycophantic outputs, repeating or agreeing with user inputs even when they are inaccurate.	These issues arise due to limitations in the model's training objectives as well as the model's system prompt and the lack of an internal representation of truth, leading to inconsistent and unreliable performance.
Unpredictable and emergent outputs	Generative AI models, especially those trained on large-scale datasets, often display outputs or capabilities not explicitly programmed for or anticipated during development.	Unpredictability results from complex interactions between model architecture, training data, and task formulation.
Hidden and systemic bias in data and outputs	Training data used in generative AI models can contain statistical irregularities or imbalances reflecting societal, institutional, or operational prejudices.	Because generative AI models learn correlations rather than causal structures, they may reinforce data biases during inference, leading to skewed outcomes.
Brittleness	Generative AI models may perform poorly when exposed to inputs that deviate from what was included in the training data or even when exposed to only relevant data.	This issue stems from data that is insufficiently diverse, relevant, or reliable; and is also noisy, corrupted, or incomplete.
Opacity and lack of explainability	The process through which certain inputs are converted into outputs is not readily interpretable by humans.	Due to a lack of pre-defined rules in ML systems, and convoluted layers of parameters in deep neural networks.
Weak anchoring of human values	Generative AI systems do not hold human values or ethical judgement.	Human values are not natively represented in transformer models; they are imposed externally through curation, fine-tuning, and reinforcement, and the result remains partial and contestable.

[2] Table adapted from: Global Commission on Responsible AI in the Military Domain. (2025). *Responsible by design: Strategic guidance report.* The Hague Centre for Strategic Studies.

Because generative AI is able to deliver outputs in fluent, natural and conveying language, users, including researchers, may be inclined to accept that information without critical scrutiny. This risk is compounded by researchers' biases, which can further shape how they interpret and process AI-generated outputs. In interpreting these results researchers should thus not only be aware of the technical limitations of generative AI, but also of their own biases. This is achieved by actively aiming to maintain critical thinking and by practicing verification and precision. It is necessary to always consult original sources, secondary literature, and empirical data combined with subject matter expertise to cross-verify information generated by generative AI tools.

Implications

Writing & analysis

- Always triangulate AI-generated outputs with original sources, secondary literature and empirical data while conducting research.
- Consider embedded biases in AI-generated literature reviews or draft texts caused by training data and prompt design.
- Push back on AI-outputs that do not question your existing assumptions. Agreement is not validation.

Data science & coding

- Be sceptical about agent-generated code. A working script does not mean correct output.
- Always validate pipelines manually: test against edge-cases and samples with a known ground truth if available.

2.2. Practice Active Learning

As generative AI affords researchers access to vast amounts of information, it should be used as a tool that supports learning and creativity, rather than as replacing these human facets by incentivising cognitive offloading. Researchers should always be critical about how to integrate generative AI into any research outputs. Researchers should focus more on quality over quantity, especially as generative AI content becomes more widespread, which will at first inevitably lead to the production of more rather than less information. Researchers can learn from using generative AI as a research tool, both methodologically and substantively. This, however, needs to be an active learning process, which researchers consciously engage in. Importantly, institutes should encourage researchers to take the time to do this, as in the long run, this will make researchers more effective.[3]

[3] A useful heuristic for structuring reflective practice is the traffic light model, which distinguishes between tasks for which generative AI should not be used at all (red), tasks where it can serve as augmentation under active human oversight (orange), and tasks where its use can be more straightforwardly delegated (green). The boundaries between these categories are not fixed; they depend on the research stage, the sensitivity of the material, and the stakes of the output. Even tasks that initially appear "green" may, on closer inspection, warrant more caution. Applying this model is itself an act of reflective learning, and it benefits from being discussed within teams rather than determined individually.

Table 3. Examples of questions to ask before, during, and after working with a large language model

Questions to ask before starting work with generative AI	Questions to ask while working with generative AI	Questions to ask after working with generative AI
Why do I want to use generative AI? What are the concrete goals that I want to achieve or tasks that I want to complete with the use of generative AI?	What am I learning from generated outputs? Am I actively engaged with generated outputs or am I passively prompting?	What support, if any, is missing in my work using generative AI? How has the model helped me or prevented me from achieving the goals or completing the tasks that I intended?

Implications

Writing & analysis

- Identify why you engage with generative AI in your projects. Be clear about what function it serves in, for example, research design, literature review or writing.
- Reflect on the use of AI: identify both how it improved and impeded your personal development and research workflow.

Data science & coding

- For every agent, you have to ask yourself: what is it exactly that I want it to build? Does the agent understand my intentions precisely, or does it just pretend that it does?
- Consciously engage in the process. Why does the agent want to use this method for this problem – and do I agree with its reasoning?

2.3. Consider Fit Between Tool and Purpose

Recognising that no single generative AI tool is universally superior is fundamental to sound practice. A rich and growing ecosystem of tools and applications exists, each with distinct strengths, limitations, and specialisations reflective of the scope of potential tasks and inputs. Approaching generative AI with intentionality and fluency therefore also means actively exploring this variety, rather than habitually defaulting to one option. An overreliance on a single generative AI model can simultaneously erode skills while also limiting researchers' ability to adjust and update their workflows in response to new technical developments. Therefore, it is equally important to maintain a calibrated perspective when evaluating performance: an application cannot be judged by one misstep, nor can infallibility be assumed from a single victory. Conversely, staying informed is also an ongoing responsibility. Generative AI technologies will continue to evolve rapidly, and researchers should remain aware of new tools, emerging capabilities, and best practices through professional networks, peer exchange, and direct experimentation.

Implications

Writing & analysis

- Keep up-to-date with model developments, emerging capabilities and best practices.
- Avoid overreliance on one model: make it a habit to try out different models once in a while.

Data science & coding

- Not everything should be delegated to agents – and not every agent will do a good job.
- In big data workflows, match the task to the model: simple tasks can be done quicker and more efficiently by smaller models, while complex reasoning or planning tasks will need the best models.
- Avoid a vendor lock-in: regularly use tools from multiple providers.

2.4. **Safeguard Integrity**

While generative AI can contribute to many stages of a research workflow, the choice whether or not to use it needs to be deliberate. As such, for its integration into specific research stages, there needs to be an understanding of its specific benefits. For this to be done productively and consistently, it is essential to stay up to date with best and worst or good and bad practices, and make adjustments accordingly.

Ideally, before deploying products with a generative AI component, a light touch risk assessment should be conducted, to verify the robustness and safety of the tooling, and ensure that it adheres to data governance standards. It also involves an assessment of generative AI providers and their policies, ensuring that the systems are transparent, ethical, and comply with all relevant regulations, including the European Union's AI Act stipulations for high-risk and general-purpose AI systems, as well as with the recommendations of the European Commission's Living Guidelines on the Responsible Use of Generative AI in Research, which set out the European framework for responsible AI use by researchers and research organisations.[4] Researchers should be aware of the AI's capabilities and limitations, including how the system engages with users, makes decisions, and handles data. Ultimately, these considerations will help ensure that the tools that knowledge institutes use align with and contribute to their missions.

Implications

Writing & analysis

- Consider whether the use of generative AI actually enhances the analytical rigour of your research and argumentation, and identify potential data security and confidentiality risks associated with its use.
- Default to institutionally approved tools (such as HCSS internal AI tool) for any work involving confidential, proprietary or non-public material.

Data science & coding

- Conduct a risk-assessment before deployment of a pipeline that involves generative AI: what *could* go wrong? What would the probability and impact be?
- Ensure reproducibility of the results. Use version-control for models and prompts.

[4] European Commission, Directorate-General for Research and Innovation. (2026). *Living guidelines on the responsible use of generative AI in research* (ERA Forum Stakeholders' Document, 3rd version)

2.5. **Maintain Confidentiality**

Beyond being aware of the back-end limitations of generative AI it is also important to be safe in the types of information that are shared with these systems as well as how these systems are integrated into and interact with institutional information environments. There are multiple security risks that can be introduced with the use of generative AI models, and therefore it is important to both practice safe communication and to create safe environments. This often requires implementing technical safeguards as generative AI becomes integrated within institutional systems. Unless indicated otherwise, interactions with generative AI tools may be stored or processed, and in most cases used as training data for model improvement. Sharing personal data[5] or any work-related information that can be traced back to an identifiable person is regulated by the European Union's General Data Protection Regulation (GDPR). Uploading this data to an online generative AI tool may lack a lawful basis to do so and may therefore violate the GDPR. It may also run counter to confidentiality conditions that are part of contract research undertaken on behalf of research funders. Data that falls under either of these two categories should therefore not be uploaded to an online generative AI tool.

Following these considerations of security and price-quality correspondence, institutes can broadly consider four deployment options for using generative AI. The most accessible and least secure option is to take a subscription with a generative AI provider such as OpenAI (ChatGPT), Anthropic (Claude), or Google (Gemini). These providers store and process all data used on their own servers, including uploaded documents, chat history and web searches, leaving users no control over where their data goes. A second option is to use a cloud provider separate from the generative AI provider, such as Microsoft Azure or Amazon Web Services, which can host models on servers around the globe and often allow users to select data centres in particular countries or regions. In this case chat history is stored locally or with a separate provider, and the model only processes requests rather than storing data. Utilising a region-specific sovereign cloud builds on this concept, but limits deployment to dedicated, smaller-scale cloud providers that are running their own (open-source) generative AI-models. These providers often offer more control (such as the physical location of your data storage) and better fit region-specific needs. For European knowledge institutes like HCSS, a key difference is that they are based in the European Union (EU), meaning they are not exposed to non-EU legislation such as the CLOUD act that can potentially violate the European GDPR.[6] The fourth and most secure option is to host generative AI models entirely locally on institutionally owned infrastructure, fully ensuring that data is never kept or processed externally. This is also the most cost-intensive option in terms of infrastructure and maintenance, especially when model providers keep offering non-profitable prices to grow their customer base.

---

5 'Personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. From: *General Data Protection Regulation, Art. 4 – Definitions.* (n.d.). GDPR-info.

6 The CLOUD Act (2018) is U.S. legislation allowing their law enforcement to demand companies (i.e. Microsoft) to hand over data stored on foreign servers, including the EU. This law is in direct conflict with the European privacy law (GDPR). See: Clarifying Lawful Overseas Use of Data Act, H.R. 4943, 115th Cong. (2018).

Table 4. Four options for organisations to access generative AI models

Deployment type	Security level	Price-Quality	Data residency
Generative AI-model Provider	Low	Maximum	Provider servers
Global Cloud	Moderate	High	Selected region
Sovereign Cloud	High	Moderate	Local country/EU servers
Local infrastructure	Maximum	Low	Internal

Considerations of confidentiality and data security increasingly intersect with another dimension: the strategic dependencies that can follow from sustained use of a particular provider's infrastructure. The choice of an AI-model provider is not only a question of security level and price-quality balance; it can also be understood as a choice about where data, workflows, and ultimately institutional capabilities are anchored. For European knowledge institutes, this raises questions about long-term reliance on non-European providers, questions that, in relation to generative AI, have until recently received less attention than they have for other forms of digital infrastructure.

This consideration does not point to a single course of action, and the most appropriate choice will vary by project, sensitivity, and available resources. It does suggest, however, that the deployment options described above can be weighed not only against immediate operational criteria but also against the longer-term implications of where institutional research practice is hosted. Where projects allow, giving meaningful consideration to European or sovereign cloud options is one way to take this dimension into account.

Implications

Writing & analysis

- Never enter sensitive, classified or personal data into externally hosted generative AI-models that do not guarantee safe data residency or GDPR compliance.[7]
- Default to institutionally approved tools (such as HCSS internal AI tool) for any work involving confidential, proprietary or non-public material.

Data science & coding

- Never pass sensitive, classified or personal data to externally hosted generative AI-models that do not guarantee safe data residency or GDPR compliance.
- When using coding agents on real datasets, use anonymised or synthetic data samples during development.

[7] Personal data as understood under the GDPR and Dutch AVG. Classified information refers to information shared under NDA or classified documents shared by government clients (in the Dutch context Departementaal Vertrouwelijk and upwards). Sensitive information includes proprietary or non-public materials generated by HCSS or clients.

2.6. Use Deliberate Prompting

Before using generative AI, it is important to consider whether it is truly the right tool for the task, or whether other approaches, such as for example the use of Boolean operators, might be equally or more effective. When generative AI is the appropriate choice, the quality of its outputs depends heavily on the quality of the prompts used to guide it. This begins with a clear sense of purpose: knowing what you want to achieve and what you expect to get out of a given prompt. From there, advanced prompt engineering techniques can significantly enhance the reliability of responses. Prompt engineering includes providing detailed instructions, breaking complex tasks into manageable sub-tasks, and requesting justifications or reasoned explanations. Generally, prompts should use short sentences with clear and correct grammatical structure, avoid vague directionals (this, that, etc.), and be explicit about the content that is being referenced. In some instances, starting a new conversation may also offer better results. Prompts should also be tailored to reflect the desired end goal and the specific research stage at which they are being employed. Throughout the workflow, they should be regularly reviewed and refined to ensure they remain aligned with these research intentions. Ensuring expert prompting based on contextualised knowledge is critically important, as appropriate use of generative AI relies heavily on well-informed inputs that take into account the specific nuances and intricacies of any particular subject matter.

Implications

Writing & analysis

- Identify what you seek to achieve and what you expect to get out of a given prompt.
- Provide detailed instructions, breakdown complex tasks and request justifications.
- Adopt prompts to reflect the specific research stage for which the AI-tool is being used, be explicit about this.

Data science & coding

- Use elicitation prompting while planning a project.
- Make sure to structure sessions around the CRISP-DM phases: scope first, explore the data second and built only once you have clear overview.[8]

2.7. Ensure Transparency

When generative AI is used as part of a research process, it should be reported in a transparent and proactive manner. Within project teams, researchers should communicate with one another how they are employing generative AI. Simultaneously, there needs to be awareness of the fact that the opaque nature of generative AI outputs may sometimes make it challenging to attribute credit for specific ideas or concepts to original authors. Therefore, the use of generative AI to identify insights should always be complemented with consultation of primary or secondary sources, including through the use of Retrieval Augmented Generation Techniques outside of the generative AI tool, and through independent human assessment.

[8] CRISP-DM (Cross-Industry Standard Process for Data Mining) is the standard approach to working with big data. It ensures a responsible, holistic way of working with data. From: Hotz, N. (2018, September 10). *What is CRISP DM?* Data Science PM.

Additionally, HCSS makes use of AI-labels to transparently indicate the use of AI in its data analysis and knowledge production. Based on initiatives such as the AI Assessment Scale (AIAS), DISCLOSE-AI, and MMMLABEL,[9] HCSS adopts the following labels to indicate the use of AI in its activities:

Table 5. Four levels of AI involvement in research activities

Label	Description	Role of AI	Role of researcher
No AI	Traditional analytical work in which no generative or agentic AI tools are used.	None.	Research, analysis, writing, review and synthesis are carried out entirely by the human researcher.
AI-supported	Data analysis and knowledge production in which the human researcher remains central but uses AI as a supporting tool.	Assists with brainstorming, structuring ideas, reviewing arguments, identifying blind spots, or improving clarity.	Core analytical tasks such as literature review, conceptual development, substantive judgement and final writing remain human led.
AI-driven	AI performs substantial parts of the data analysis and knowledge production process.	Drafts text, conducts extensive literature reviews, summarises data sources, generates analytical options or produces first versions of deliverables.	Retains responsibility for direction, interpretation, validation and final result; role shifts from primary producer to supervisor, editor and quality assurance.
AI-orchestrated	A researcher designs, configures and monitors an AI-based system, after which the system operates with a high degree of autonomy.	Agentic workflows in which AI-systems collect information, analyse data, generate outputs and make procedural decisions with limited human intervention.	Responsibility shifts from individual task execution to system design, governance, monitoring and accountability.

Implications

Writing & analysis

- Be transparent about the use of AI in different research phases and activities. Keep a record of AI-tools used during these stages.
- Transparently communicate the use of AI to colleagues, funders, and research consumers, utilising the AI-labelling framework presented above.

Data science & coding

- Make it clear in the output what texts are human-written and what is AI-generated. In the documentation of research and tools, include an overview of which AI models are used along which steps.

[9] Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). *The AI Assessment Scale* ;Me & Machine. (n.d.). *Me & My Machine (MMM) labels*.

2.8. **Accommodate Experimentation**

To maximise user proficiency with generative AI, it is essential to engage in experimentation, learn best practices from peers, and, most of all, exercise patience. Generative AI is continuously evolving, presenting new features and capabilities over time. This journey is far from static and where it leads next remains to be seen. It requires continuing discussion about the conditions under which generative AI can be used effectively and responsibly. As such, researchers should try things that may fall outside of the direct scope of a particular research project in order to further familiarise themselves with various tools and functionalities. Learning can further be facilitated by fostering regular dialogues and feedback loops, encouraging open communication, and involving experts from different fields to cross-verify, refine, and enhance outputs.

Experimentation tends to be more productive when it is ringfenced: dedicated environments or sandbox configurations make it easier to try out tools and workflows without exposing live projects or sensitive data. Within HCSS this includes setting up small-scale test environments, including ones using locally hosted or open-source models, in which researchers can explore what does and does not work without operational consequences. Sharing the results of such experiments, including those that did not yield the expected outcomes, can be at least as valuable as sharing successful cases.

Implications

Writing & analysis

- Create safe environments (for example in HCSS's internal AI tool) to experiment with the use of AI in different stages of research.
- Share effective workflows with other colleagues to further institutional learning and AI-adoption.

Data science & coding

- Set up sandbox environments to test with new models or methods.
- Dedicate some time to explore new agent capabilities or tools.
- Make sure to share workflows that worked well (or that failed) with the team, so we can learn from one another.

The Hague Centre
for Strategic Studies

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies
Email: info@hcss.nl
Website: www.hcss.nl