AI and Ethics (2025) 5:6371–6377
https://doi.org/10.1007/s43681-025-00768-8

RESEARCH

Core principles of responsible generative AI usage in research

Tim-Dorian Knöchel[1,2] · Konrad J. Schweizer[3] · Oguz A. Acar[4] · Atakan M. Akil[5,6] · Ali H. Al-Hoorie[7] · Florian Buehler[8] · Mahmoud M. Elsherif[9,10] · Alice Giannini[11] · Evelien Heyselaar[12] · Mohammad Hosseini[13] · Vinodh Ilangovan[14] · Marton Kovacs[5,15,16] · Zhicheng Lin[17,18] · Meng Liu[19] · Anco Peeters[2,20] · Don van Ravenzwaaij[21] · Marek A. Vranka[22] · Yuki Yamada[23] · Yu-Fang Yang[24] · Balazs Aczel[5]

Received: 27 February 2025 / Accepted: 29 May 2025 / Published online: 14 October 2025

Abstract
In a rapidly evolving Generative Artificial Intelligence (GenAI) landscape, researchers, policymakers, and publishers have to continuously redefine responsible research practices. To ensure guidance of GenAI use in research, core principles that remain stable despite technological advancement are needed. This article defines a list of principles guiding the responsible use of GenAI in research, regardless of use case and GenAI technology employed. To define this framework, we conducted an anonymised Delphi consensus procedure comprising a panel of 16 international and multidisciplinary experts in AI, social sciences, law, ethics, and scientific publishing. After three rounds of independent rating and feedback, the panel reached consensus on eight sequentially ordered principles required for responsible GenAI usage: Regulations, Data Security, Quality Control, Originality, Bias Mitigation, Accountability, Transparency, and Broader Impact. For the clear reporting of adherence to these principles, we created a checklist allowing active implementation into the research process. With these efforts, we aim to guide everyday research, support the development of further specified regulations, policies, and guidelines, and promote discussion about GenAI use in research.

Keywords Ethics · Generative artificial intelligence · Guidelines · Policy and regulation · Research integrity

Generative artificial intelligence (GenAI) refers to AI systems capable of producing novel content (e.g., text, images, code, or data) in response to input prompts. Powered by large language models and generative algorithms, GenAI tools have a growing impact on how research is conducted [1]. While GenAI can drastically speed up certain tasks, its use presents serious risks for research integrity, security, and may diffuse responsibility for adverse outcomes [2]. Ongoing efforts by policymakers, scientists, publishers, and research institutions aim to address these risks by formulating regulations for the responsible use of AI in research. Such efforts apply at varying scopes, from broad ethics frameworks to regional research and development guidelines, academic consensus frameworks, publication ethics standards, down to publisher and journal policies, and, finally, discipline-specific or tool-specific checklists. Some approaches refer to all forms of AI usage, while others target GenAI in particular, and further distinctions can be made on whether the regulation is descriptive or prescriptive. Although these regulatory efforts are vital in guiding research practices, the efficacy and sustainability of any regulation are significantly challenged by the continuous evolution of GenAI. The resulting sudden and unforeseen developments can disorient researchers, causing confusion about appropriate and responsible scientific practices when using GenAI. Therefore, a broad and enduring foundation for the use of GenAI in research is required, serving as a guide for regulatory efforts alongside evolving AI technologies. Accordingly, this article defines overarching principles within a framework to guide the responsible use of GenAI in research, regardless of the use case or employed model. We also provide essential preparatory steps and offer a comprehensive checklist to facilitate adherence to these principles.

Tim-Dorian Knöchel and Konrad J. Schweizer shared first authorship.

Extended author information available on the last page of the article

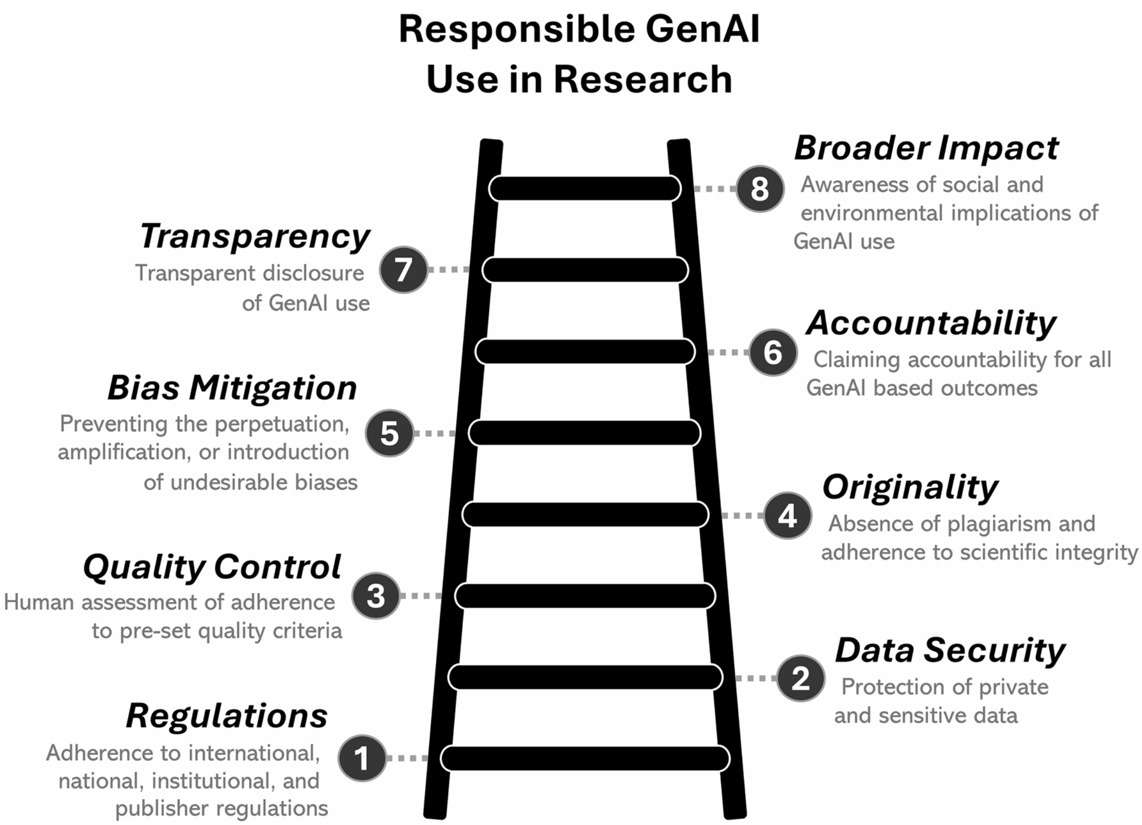

Fig. 1 Core principles of responsible GenAI usage in research

While this framework results from a descriptive academic consensus on GenAI, with issues, such as content originality and hallucinations being specific to it, many principles also guide the responsible use of other AI tools. To identify principles, we conducted a Delphi consensus procedure comprising a panel of 16 international and multidisciplinary experts in AI, social sciences, law, ethics, and scientific publishing. The procedure was preregistered on OSF [https://doi.org/10.17605/OSF.IO/R4W9B]. All methodological details can be found in the supplementary text of this article [https://osf.io/uvtx5]. Consensus was reached for eight principles (Fig. 1). All principles are organised in a sequential order, starting with the most general requirements, which should be addressed first, followed by steps that are relevant only if the previous principles are satisfied. The framework distinguishes itself from higher-level ethics codes by translating broad scientific and societal values into concrete action recommendations for scientific use [3] and complements more narrow publishing and discipline-specific guidelines by defining overarching principles through an expert committee.

1 Regulations

Researchers must follow regulations, policies, and guidelines applicable at institutional, national, and international levels [4], as well as those set by publishers regarding their GenAI use. Ethical risk assessment and compliance with the ethical review board's expectations help further mitigate ethical risks. In international and multi-institutional projects, differing AI regulations and ethical guidelines may apply. Ensuring that team members comply with the rules relevant to them throughout the research is critical.

2 Data security

Researchers should make all reasonable efforts to ensure the security of private and sensitive data. Entering identifiable data into GenAI systems involves the risk of unauthorised third-party access, inadvertently compromising research integrity. To prevent compromising private or sensitive data, researchers can employ different strategies, such as using anonymous or pseudonymised identifiable information. Equally, identifying AI providers with more robust privacy policies and consulting data protection teams (external or institutional) is advised.

3 Quality control

Like human contributions, GenAI outputs should follow quality standards to ensure good scientific practice. GenAI contributions cannot be blindly trusted, and require human verification concerning accuracy and precision (e.g., *is it correct?*); logical reasoning (e.g., *is it consistent and coherent?*); relevance (e.g., *is it topical, comprehensive, and up-to-date?*); and context-specific criteria, such as evidence standards (*professional quality*). Researchers should pre-specify procedures to verify these criteria and involve at least one human contributor to assess GenAI contributions and outputs based on the specified criteria. These criteria and verification procedures should be documented and transparently reported in any associated work.

4 Originality

When using GenAI, researchers should ensure that all research components, including text in the main and supplementary documents, figures, data, and metadata, are free from plagiarism and accurately reference original sources. For instance, GenAI systems sometimes fail to provide accurate references, risking the misrepresentation of existing research. Therefore, human authors must check the originality of GenAI outputs and ensure proper acknowledgement of used sources [5]. Avoiding the direct use of GenAI outputs in publications is a cardinal way of preventing originality issues.

5 Bias mitigation

Researchers should make all reasonable efforts to avoid perpetuating, amplifying, or introducing undesirable biases when using GenAI (e.g., existing gender and racial bias [6]). Uncritical reliance on GenAI can reinforce societal or academic power structures, stereotypes, or biased consensus. Researchers can employ various strategies to assess or mitigate AI biases [7]. They can consult previous evaluations of used models, or follow checklists during the implementation or dissemination of the GenAI outputs [8]. A general strategy for bias mitigation can be vetting the research using available bias benchmarks within the used AI models, domain experts, and diverse perspectives.

6 Accountability

Accountability for one's scientific work is among the hallmarks of good science and facilitates society's trust in research results. In all published content, only humans remain accountable for the strengths and weaknesses of presented work. Unlike humans, AI systems do not make conscious decisions; they are not liable agents, and, therefore, they cannot be held accountable or sanctioned for any of their errors. Researchers who use GenAI systems should ensure that they only use models in contexts where they have sufficient expertise and information to evaluate the model's output [9].

7 Transparency

When using GenAI for research purposes, it is imperative to clearly document and communicate GenAI contributions and their validation process by humans. Acknowledging and reporting the use of GenAI tools promotes accountability, fosters trust, and facilitates verification and replication. GenAI tools are constantly being updated, but changes may not necessarily be reflected in the model version, so providing dates of usage is good practice [10]. Output may also be sensitive to the prompts [11], making their documentation informative for replication. Due to response stochasticity and iterative involvement, a complete documentation of GenAI usage may be cumbersome, and certain use cases (e.g., copy-editing) might not require detailed reporting. Field or topic-specific guidelines may be needed to ensure consistency [3].

8 Broader impact

It is crucial that scientists are aware of the potential social and environmental impacts of using AI [12]. Since the training and development of AI consumes substantial energy, it produces considerable emissions [13]. When using AI, researchers should consciously consider its energy consumption. It is important to question the energy efficiency of a deployed model and to seek out more efficient options if they are able to provide comparable results. Furthermore, as GenAI will likely replace more and more areas of scientific work (e.g., data analysis, programming), researchers should pay attention to how it affects the development of their own scientific skills [14]. In addition, employing GenAI can displace or limit the involvement opportunities of co-workers, further increasing social inequalities.

Table 1 Principles, Preparation, and Checklist Questions for Responsible GenAI Usage in Research

Principle	Preparation	Checklist
Regulations	Ensure that your GenAI usage and application follow the applicable institutional, national, international, and publishers' regulations and policies. In addition, identify and address potential ethical risks associated with your GenAI application in accordance with your ethical review board	Do the applicable regulations and policies allow for the use of GenAI tools in research in your work?
Data Security	To prevent compromising private or sensitive information when working with GenAI, you should (i) anonymise or pseudonymise all input data, (ii) opt out of data usage and storage in the GenAI application(s) where possible, and (iii) consult an expert on the topic of data security if needed. Whenever relevant, inform the participants how their data will be used in the GenAI application	Is your research using GenAI compatible with data privacy and security regulations?
Quality Control	Quality Criteria Have predefined verifiable criteria for correctness, reasoning, relevance, and professional quality Quality Checks A human contributor should assess whether the GenAI-generated outputs or their modified version satisfy previously set quality criteria *Correctness:* Is the output accurate and correct according to the set quality criteria? *Reasoning:* Does the output make logical sense? Is it free of contradictions? *Relevance:* Is the output relevant, comprehensive, and up to date according to human expertise? *Professional quality:* Does the output satisfy all the other set quality expectations?	Do you have specific quality criteria for the outputs of your GenAI tool? Did a human contributor assess the outcomes by the set quality criteria and checks?
Originality	Ensure that all components of the work, including text, data, figures, and images, are free from plagiarism in contexts where concerns of plagiarism apply. Make sure that the work of other parties is fully acknowledged and used in compliance with academic integrity	Did you make sure not to use someone else's work without appropriate accreditation?
Bias Mitigation	AI can create random or systematic biases. Ensure you systematically check whether your GenAI outputs reflect or perpetuate existing biases, power structures, stereotypes or biased consensus within society or academia, or introduce novel biases. Mitigate these biases where possible and relevant	Did you try to make sure that your project does not create, reflect or perpetuate biases or unfairness due to GenAI usage?
Accountability	Make sure that you understand the GenAI output you are using and can explain it, so that you can accept accountability for its quality, originality, and fairness	Can you explain the GenAI output? Do you accept accountability for all resulting content?
Transparency	Make sure that the process of using GenAI tools and their input to your work is documented and clearly communicated to the reader. Be clear about which tools and versions were used	Is the use of GenAI documented and reported transparently?
Broader Impact	Be aware of the environmental and societal effects of GenAI usage and consider human labour as an alternative. Also, be aware that replacing human labour affects co-workers, juniors, or researchers of marginalised backgrounds who could be involved in and profit from your research	Are you aware of the social, economic, and environmental impacts that could arise from your GenAI use and have you considered alternatives?

The checklist can be filled out via a dedicated Shiny app under [https://github.com/marton-balazs-kovacs/CorePrincipleGenAIChecklist], which has been archived with Zenodo [15]. It is advised to provide a link to the generated report within a preregistration, preprint, or article.

Checklist

To ensure the practical applicability of these eight principles, we provide concrete preparations and checks for researchers who want to use GenAI in their work. Each item in the checklist corresponds to one of the eight principles outlined above, translating them into concrete decision points for researchers (see Table 1). It displays a framework that guides users in deciding whether a given GenAI application can be used for research purposes. This scheme includes (i) a comprehensive description of the fundamental value meant to be satisfied (principle), (ii) a short description of the steps that can be taken to satisfy the principle (preparation), and (iii) the checklist questions must be answered with '*yes*' or sufficiently explained when answered with '*no*', to consider the checklist to be completed. The checklist is intended to complement and support, rather than replace, a user's own critical attitude towards responsible GenAI usage.

A collection of AI use guidance models, and their comparison is included in the supplement.

Conclusion

Awareness of these eight principles contributes to responsible GenAI use on both a general and concrete level. They serve as an initial take to achieve a stable guide in an ever-changing AI landscape and inform the formulation of further guidelines concerning ongoing AI developments in research.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43681-025-00768-8.

Author contributions **Conceptualisation**: T.-D.K., K.J.S., O.A.A., A.M.A., A.H.A.-H., F.B., M.E., A.G., E.H., M.H., V.I., Z.L., M.L., A.P., D.v.R., M.A.V., Y.Y., Y.-F.Y., and B.A. **Data curation**: T.-D.K. and K.J.S. **Formal analysis**: T.-D.K. and K.J.S. **Investigation**: T.-D.K. and K.J.S. **Methodology**: T.-D.K., K.J.S., and B.A. **Project administration**: T.-D.K., K.J.S., and B.A. **Software**: M.K. **Supervision**: T.-D.K. and B.A. **Validation**: T.-D.K., K.J.S., O.A.A., A.M.A., A.H.A.-H., F.B., M.E., A.G., E.H., M.H., V.I., Z.L., M.L., A.P., D.v.R., M.A.V., Y.Y., Y.-F.Y., and B.A. **Visualisation**: T.-D.K., K.J.S., and M.K. **Writing - original draft**: T.-D.K., K.J.S., O.A.A., A.M.A., A.H.A.-H., F.B., M.E., A.G., E.H., M.H., V.I., Z.L., M.L., A.P., D.v.R., M.A.V., Y.Y., Y.-F.Y., and B.A. **Writing - review & editing**: T.-D.K., K.J.S., O.A.A., A.M.A., A.H.A.-H., F.B., M.E., A.G., E.H., M.H., V.I., M.K., Z.L., M.L., A.P., D.v.R., M.A.V., Y.Y., Y.-F.Y., and B.A.

Funding For all authors with one exception, no funding was received to assist with the preparation of this manuscript.Mohammad Hosseini is supported by the National Institutes of Health (NIH) National Center for Advancing Translational Sciences (UM1TR005121) and the NIH Office of Data Science Strategy/Office of the NIH Director pursuant to OTA-21-009, "Generalist Repository Ecosystem Initiative (GREI)" through Other Transactions Agreement (OTA) Number OT2DB000013-01. The views and opinions expressed here do not represent the views and opinions of the NIH or US government.

Data availability All anonymised raw and processed data, as well as the survey materials, are publicly shared on the Open Science Framework page of this project [https://doi.org/10.17605/OSF.IO/T3SFB]. Our methodology and data analysis plan were preregistered before the project. The preregistration document can be accessed at [https://doi.org/10.17605/OSF.IO/R4W9B].

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethical Approval Since all experts who participated in the Delphi consensus procedure are members of the author team, they are not considered participants. Consequently, informed consent from them was deemed unnecessary.

References

1. Kwon, D.: Is it OK for AI to write science papers? Nature survey shows researchers are split. Nature **641**, 574–578 (2025)
2. Hanson, B., et al.: Garbage in, garbage out: mitigating risks and maximizing benefits of AI in research. Nature **623**, 28–31 (2023)
3. Lin, Z.: Beyond principlism: practical strategies for ethical AI use in research practices. AI Eth **5**, 2719–2731 (2025). https://doi.org/10.1007/s43681-024-00585-5
4. European Parliament. EU AI Act: first regulation on artificial intelligence. *Topics| European Parliament*. https://www.europa1-europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (2023)
5. Dien, J.: Editorial: generative artificial intelligence as a plagiarism problem. Biol. Psychol. **181**, 108621 (2023)
6. Birhane A, Dehdashtian S, Prabhu V, Boddeti V (2024) The dwark side of dataset scaling: evaluating racial classification in multimodal models. In Proceedings of the 2024 ACM conference on fairness, accountability, and transparency pp. 1229–1244 (Association for Computing Machinery, New York, NY, USA, 2024). https://doi.org/10.1145/3630106.3658968.
7. Agarwal, A., Agarwal, H., Agarwal, N.: Fairness score and process standardization: framework for fairness certification in artificial intelligence systems. AI Eth **3**, 267–279 (2023)
8. Nazer, L.H., et al.: Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit. Health **2**, e0000278 (2023)
9. Hosseini, M., Resnik, D.B., Kristi, H.: The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. AI Eth (2024). https://doi.org/10.1007/s43681-024-00493-8
10. Hosseini, M., Resnik, D.B., Holmes, K.: The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. Res. Eth **19**, 449–465 (2023)
11. Wang, L., et al.: Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. Npj Digit. Med. **7**, 1–9 (2024)
12. Bolte, L., van Wynsberghe, A.: Sustainable AI and the third wave of AI ethics: a structural turn. AI Ethics (2024). https://doi.org/10.1007/s43681-024-00522-6
13. Strubell E, Ganesh A, McCallum, A. (2020) Energy and policy considerations for modern deep learning research. In: Proc. AAAI Conf. Artif. Intell.**34**, 13693–13696
14. Heersmink, R.: Use of large language models might affect our cognitive skills. Nat. Hum. Behav. **8**, 805–806 (2024)
15. Kovacs, M., Knöchel, T.-D., Schweizer, K. J., & Aczel, B. (2025). marton-balazs-kovacs/CorePrincipleGenAIChecklist: Core Principles of Responsible Generative AI Usage in Research Checklist (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.15657712

Authors and Affiliations

Tim-Dorian Knöchel[1,2] · Konrad J. Schweizer[3] · Oguz A. Acar[4] · Atakan M. Akil[5,6] · Ali H. Al-Hoorie[7] · Florian Buehler[8] · Mahmoud M. Elsherif[9,10] · Alice Giannini[11] · Evelien Heyselaar[12] · Mohammad Hosseini[13] · Vinodh Ilangovan[14] · Marton Kovacs[5,15,16] · Zhicheng Lin[17,18] · Meng Liu[19] · Anco Peeters[2,20] · Don van Ravenzwaaij[21] · Marek A. Vranka[22] · Yuki Yamada[23] · Yu-Fang Yang[24] · Balazs Aczel[5]

✉ Tim-Dorian Knöchel
t.d.knochel@gmail.com

Konrad J. Schweizer
konrad.schweizer@ru.nl

Oguz A. Acar
oguz.acar@kcl.ac.uk

Atakan M. Akil
akil.atakan@pte.hu

Ali H. Al-Hoorie
hoorie_ali@hotmail.com

Florian Buehler
florian.buehler@fhv.at

Mahmoud M. Elsherif
mahmoud.medhat.elsherif@gmail.com

Alice Giannini
a.giannini@maastrichtuniversity.nl

Evelien Heyselaar
evelien.heijselaar@ru.nl

Mohammad Hosseini
mohammad.hosseini@northwestern.edu

Vinodh Ilangovan
Vinodh.Ilangovan@tib.eu

Marton Kovacs
marton.balazs.kovacs@gmail.com

Zhicheng Lin
zhichenglin@gmail.com

Meng Liu
mengliu@bfsu.edu.cn

Anco Peeters
mail@ancopeeters.com

Don van Ravenzwaaij
d.van.ravenzwaaij@rug.nl

Marek A. Vranka
vranka.marek@gmail.com

Yuki Yamada
yamadayuk@gmail.com

Yu-Fang Yang
yu-fang.yang@fu-berlin.de

Balazs Aczel
balazs.aczel@gmail.com

1 Department of Social and Cultural Psychology, Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

2 Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

3 Department of Clinical Psychology, Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

4 Department of Marketing, King's Business School, King's College London, London, United Kingdom

5 Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary

6 Institute of Psychology, University of Pécs, Pécs, Hungary

7 Jubail English Language and Preparatory Year Institute, Royal Commission for Jubail and Yanbu, Jubail, Saudi Arabia

8 Department of Business and Management, Vorarlberg University of Applied Sciences, Dornbirn, Austria

9 Department of Psychology, University of Birmingham, Birmingham, United Kingdom

10 School of Psychology and Vision Sciences, University of Leicester, Leicester, United Kingdom

11 Department of Criminal Law and Criminology, Faculty of Law, Maastricht University, Maastricht, The Netherlands

12 Department of Communication and Media, Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

13 Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

14 TIB Leibniz Information Centre for Science and Technology, Hannover, Germany

15 Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary

16 MNB Institute, John von Neumann University, Kecskemét, Hungary

17 Department of Psychology, University of Science and Technology of China, Hefei, China

18 Department of Psychology, Yonsei University, Seoul, Republic of Korea

19 School of English and International Studies, Beijing Foreign Studies University, Beijing, China

20 School of Artificial Intelligence, Radboud University, Nijmegen, The Netherlands

21 Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands

22 Department of Marketing Communication and PR, Institute of Communication Studies and Journalism, Charles University, Prague, Czech Republic

23 Division of Experimental Natural Science, Faculty of Arts and Science, Kyushu University, Fukuoka, Japan

24 Division of Experimental Psychology and Neuropsychology, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany