fteval JOURNAL
for Research and
Technology Policy Evaluation

November 2024, Vol. 55, pp. e4, 1-35
DOI:10.22163/fteval.2024.660

THE INTERSECTION OF AI AND RTI (POLICY) EVALUATION: PRINCIPLES AND CONSIDERATIONS

AUTHORS: STEFANIE SCHUERZ, CHARLOTTE D'ELLOY, MICHAEL DINGES

CONTRIBUTORS: SABINE CHAI, THOMAS PALFINGER, MICHAEL STRASSNIG, ISABELLA WAGNER, MAGDALENA WICHER
DOI: 10.22163/FTEVAL.2024.660

PREAMBLE

In May 2023, the Austrian Platform for Research and Technology Policy Evaluation (fteval) established a working group on Artificial Intelligence (AI). The working group members agreed to split into three subgroups with distinct tasks: The first subgroup would collect proprietary AI systems relevant to the evaluation phases and activities. The second subgroup would consider the evaluation system and model of interaction between its actors, as well as the possible effects that can arise from AI, and make reasonable recommendations for a strategy. The third subgroup would screen existing guidelines on the use of AI and adapt them to the fteval community, integrating good practice examples and including input and feedback from the other two groups. The final deliverables of the working group are three papers – one for each subgroup – to be shared as resources with the fteval community via the fteval journal. The working group members met between November 2023 and January 2024 to share their progress and discuss questions that had arisen during the subgroups' work. The document below is the outcome of subgroup 3 on guidelines, supplemented by information elaborated in subgroup 1 on AI systems, and intends to provide comprehensive guidance to the fteval community.

INTRODUCTION

As the digital transformation changes our society at an ever-accelerating pace, Artificial Intelligence (AI) promises to impact organisational and business processes, boost productivity, and support the analysis and visualisation of large amounts of data in an unprecedented manner. Generative AI (GenAI) systems and services show staying power beyond media hype and are in the process of establishing themselves as integral parts of research, technology, and evaluation work. At the same time, many important questions remain. Within the fteval working group on AI, we discussed various difficulties we encountered, including: How do we effectively employ AI systems while not negatively impacting the quality and rigorousness of our work? How do we (need to) disclose the specifics of our employment of AI systems in the context of our work and whom do we (need to) disclose this to? In other words, what standards of use and reporting do we need in order to integrate AI systems in our work? And how may we make informed decisions on all of these questions?

Against this background, this document is structured as follows: After an overview on the employed methodology, it provides working definitions for the specific types of AI covered in this paper and delineates the scope of AI subfields to establish a joint understanding on the multitude of AI application areas. It then outlines principles and considerations for effectively integrating AI into RTI evaluation contexts, ensuring informed decision-making and maximising the benefits of AI-driven approaches. Finally, the document describes the diverse application areas of AI within the context of research, technology and innovation (policy) evaluation, offering some insights into its multifaceted roles in daily evaluation practices of the fteval community. Annex I provides a translation of a survey sent to the fteval community, while Annex II entails a list of proprietary AI systems, including potential areas of application, pricing models, available languages, and possible alternatives.

The document focuses mostly on practical and ethical considerations of employing AI systems and solutions, not for their development and deployment. The use of AI systems for evaluating scientific research proposals and the use of AI processes for student or teacher evaluation in educational contexts are also out of scope for this paper. The main target group are evaluation practitioners working as evaluators, programme owners and managers in the Austrian RTI community.

METHODOLOGY

The present Principles and Considerations have been developed by a working group established in the context of the fteval community shortly after the release of ChatGPT-4 with the aim of exploring the impact of this highly disruptive technology on the Austrian RTI community. Split into three thematic subgroups, various aspects of this emergent cluster of technologies were explored, including a mapping of specific systems, changing roles within the RTI evaluation ecosystem, as well as existing principles and guidelines and their applicability to our RTI evaluation context. In both subgroup and plenary meetings, study design, methodologies and findings were discussed and aligned to create several documents, including the one at hand.

For these Principles and Considerations, the work plan started with a systematic literature review of existing AI guidelines from RTI, business and policy contexts. These were reviewed and core information was extracted according to a predefined set of questions to synthesise the principles and considerations laid out below. Specifically, the following information was extracted from each document:

- Country of origin, publishing organisation, year of publication
- Core focus of the text: Development or use of AI?
- What definition(s) of AI are used (explicitly and implicitly)
- Which topics on AI are addressed (e.g., trustworthiness of AI, etc.)
- Which processes and tasks are addressed?
- Which problems / problem areas are addressed?
- Which solutions are suggested?
- What topic areas are (still) missing?

Additional information was sought through existing publications to fill gaps in understanding and information. Furthermore, a short survey was developed and shared within the fteval community, to gather inputs on challenges, information gaps, attitudes, practices, as well as existing institutional guidelines regarding the use of AI within fteval member organisations. The survey employed both open and closed question formats (see Annex I) and led to a total of 20 valid responses representing Austrian RTI, RTI funding, and RTI policy organisations. Findings from this survey were used to validate and enrich the

Principles and Considerations. Finally, the first draft of the text was shared with the fteval community to review the contents and allow for adjustment. The concluding meeting of the fteval AI working group on March 15., 2024, allowed for a final discussion of the paper, which was subsequently complemented with insights from the second subgroup – mapping AI systems – to round out the text for publication.

WORKING DEFINITIONS

AI spans multiple evolving and often overlapping domains, making it difficult to identify what does and does not qualify as AI in any given context: "Most definitions of AI centre on the concept of emulating intelligent behaviour through machines, where intelligence refers to the capacity to perform complex tasks in real-world environments and learn by experience. [...] definitions of AI converge in their focus on machines or AI systems that (i) possess learning capabilities, (ii) can make intelligent decisions, (iii) influence the environment, (iv) improve tasks autonomously, and (v) exhibit human-like cognitive functions." (EC DG RTD 2023: 31)

These systems are capable of learning from data, recognise patterns, make decisions, and solve problems without explicit programming for each task. Building on a definition of the US Department of Defense, the Austrian Council for Robotics and Artificial Intelligence (Österreichische Rat für Robotik und Künstliche Intelligenz – ACRAI) characterises them as autonomous cognitive systems. They work through rule knowledge created by experts or on the basis of statistical models derived from data (machine learning, e.g. deep learning). AI systems are based upon extensive training data to analyse and identify correlations or patterns, enabling them to make predictions. This training process enhances their ability to produce improved, precise, and realistic responses, or to generate actionable insights. (ACRAI 2018)

Following Regona et al. (2022), in terms of AI research objectives, AI comprises major subfields such as machine learning, knowledge-based systems, computer vision, robotics, natural language processing, automated planning and scheduling, and optimisation. Figure 1 showcases the breadth of fields and understandings connected to AI and outlines their various components.

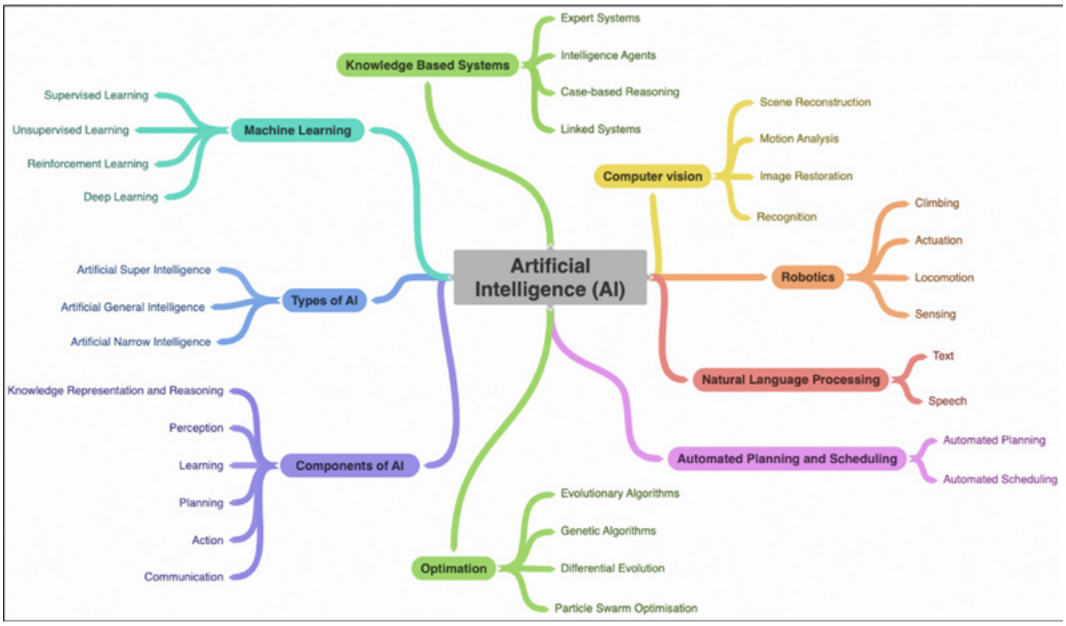

Figure 1: Components, types, and subfields of AI. (Regona et al. 2022)

Among the subfields presented above, the following are of potentially high relevance for evaluation purposes:

Machine Learning (**ML**) focuses on the development of algorithms that enable computers to learn from and make predictions or decisions based on datasets, without being explicitly programmed. Rather, they are build to receive input and predict outputs via statistical analysis, based on the respective training data. ML includes supervised learning, unsupervised learning, deep learning and reinforcement learning. (Pant 2019)

Natural Language Processing (**NLP**) is concerned with the interaction between computers and humans via "natural" human language. It involves tasks such as speech recognition, language translation, sentiment analysis, and text generation, enabling machines to interpret, generate and simulate human language. (Nancholas 2023)

Knowledge based systems (**KBS**) or **Expert Systems** (**ES**) are trained on high-level, domain-specific knowledge from human experts and are intended to mimic their decision-making abilities. They use knowledge bases, inference engines, and reasoning mechanisms to provide advice, solve problems, and make decisions. (Segreto 2019)

For this paper, we primarily focus on **Generative AI** or **GenAI**, which refers to "a type of machine learning architecture that uses AI algorithms to create novel data instances, drawing upon the patterns and relationships observed in the training data" and is thus "capable of generating seemingly new, meaningful content such as text, images, or audio from training data." (Feuerriegel et al. 2024) These new data points are generated in such a way that they "plausibly could be part of the original dataset". (Pinaya et al. 2023)

PRINCIPLES AND CONSIDERATIONS

In reviewing the relevant current literature on AI and its application in contexts adjacent to RTI, a number of principles are brought up repeatedly, intended to ensure a responsible use of AI systems that minimises harm while exhausting their potential. The following section will discuss these principles in detail.

RESPECT FOR HUMAN AUTONOMY

Various high-level guidelines on AI stress the importance of maintaining human autonomy when employing AI systems (see e.g. EC DG CNCT 2019, EC 2020, APA 2022, and OECD 2024[2019]). This means ensuring human self-determination in the sense that AI systems do not "unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans" (EC DG CNCT 2019: 12). Following Prunkl (2022), autonomy has an internal dimension, authenticity, which refers to beliefs, values and motivations that are an authentic reflection of a person's 'inner self'; and an external dimension, agency, which refers to a person's effective capacity to make and enact decisions and take charge of important aspects of their lives. For authenticity to be preserved, AI systems must not be used to manipulate or deceive people. For agency to be preserved, humans must maintain control over whether, when and how decisions concerning their lives are relegated to AI systems.

RESPONSIBILITY

When using AI systems, humans remain responsible for the accuracy, transparency, and accountability of outputs created with AI support and shared in a professional context. AI systems can serve as a valuable research tool, but they do not replace critical thinking, human expertise, and rigorous scientific methodology. Care needs to be taken to critically review the outputs created by an AI with a researcher's own expertise, to check for correctness, ethical aspects, and plausibility.

Research results derived from the analysis of AI systems should not be considered as a sole source of information for developing conclusions or recommendations. For developing these, multiple data sources, methods, and perspectives should be taken into account in order to create comprehensive and robust evaluation results. A triangulation of methods and results helps to identify inconsistencies and conflicting viewpoints and helps mitigate potential limitations of AI systems in terms of bias, limited scope of analysis, and challenges regarding explainability. Individuals should be able to understand and challenge the outcome while respecting personal data protection obligations, if relevant.

TRANSPARENCY AND EXPLAINABILITY

The Evaluation Standards for Research, Technology and Innovation Policy demand that "the evaluation process, its findings, and the subsequent recommendations are conducted and completed in a manner that is transparent and accountable for all those involved and affected." (Kohlweg 2019: 15) This also holds true when AI systems are employed in evaluation processes, meaning there is a need to be transparent about the use of AI systems and its purposes.

Transparency is particularly important when 1) AI applications are being used for analytical purposes; 2) reproducibility of results is required; and 3) the output created is likely to have an impact on decision-making processes. The OECD notes that "disclosure [of AI use] should be made with proportion to the importance of the interaction. The growing ubiquity of AI applications may influence the desirability, effectiveness, or feasibility of disclosure in some cases." (OECD 2024)

Following the OECD AI Principles, explainability pertains to "enabling people affected by the outcome of an AI system to understand how it was arrived at. This entails providing easy-to-understand information to people affected by an AI system's outcome that can enable those adversely affected to challenge the outcome, notably – to the extent practicable – the factors and logic that led to an outcome." (OECD 2024)

Tacit and explicit standards for clarifying which systems were used, for what purposes and at which point of the process will most likely be negotiated and settled in the upcoming years, depending on which systems and AI use cases are evolving. The following measures might be taken to ensure transparency in the evaluation process:

- Clearly state in the methods section which AI systems have been used for which purposes of the evaluation. Specify the respective system and relevant parameters used.
- When using AI systems for analytical purposes, we suggest keeping a record of AI-generated outputs including prompts, parameters, and outputs created by the system. Keeping a record helps to revisit the analytical process and allows for future revisions.

PREVENTION OF HARM

The fteval Standards prescribe that "activities undertaken in connection with evaluations are carried out in a manner that is ethically responsible, gender aware, and with openness toward social and cultural diversity (e.g. age, background, language)." and that "all those involved in or affected by an evaluation are treated with respect and fairness." (Kohlweg 2019: 15) Similarly, the EU-HLEG on AI's[1] Ethics Guidelines for Trustworthy AI state that "AI systems should neither cause nor exacerbate [collective or individual] harm or otherwise adversely affect human beings. This entails the protection of human dignity as well as mental and physical integrity." (EC DG CNCT 2019: 12) As such, with regard to the application and development of all research tools and methods, the responsibility for ethical development and ethical use of AI models lies both with the developer and user. When training AI models with training data for specific evaluative purposes, but also when prompting and reviewing the outputs generated by AI models, special care must be taken to identify, minimise, actively counteract and be transparent regarding the biases inherent in the training data and subsequent outputs. This is especially true when AI is used for decision-making processes, which in turn means that human verification is a must for central activities such as writing manuscripts and data scripts, peer review, proposal evaluation, and so on.

Potential discriminatory or unfair outcomes must be actively addressed and counteracted to ensure human, environmental and ecosystem safety and security. Some guidance for developers and decision-makers is provided e.g. by the UNESCO Recommendation on the Ethics of Artificial Intelligence (UNESCO 2021a) and its companion document Ethical impact assessment: a tool of the Recommendation on the ethics of artificial intelligence (UNESCO 2021b). Furthermore, in its report on Regulating AI in the UK, the Ada Lovelace Institute (2023) identifies the need for regulatory frameworks that, among others,

1 The independent High-Level Expert Group on Artificial Intelligence was set up by the European Commission in June 2018.

contain legal rights and protections for people, establish routes to redress, and introduce reporting requirements.

In addition, technical security checks need to be employed at relevant points within an AI life-cycle to address security vulnerabilities in AI models or codes developed with AI systems, in order to prevent models from manipulation, data breaches, unauthorised access, hidden malware, etc. Data collected and used by AI systems need to be handled responsibly, adhering to relevant data privacy regulations and user consent.

FAIRNESS AND NON-DISCRIMINATION

All existing AI guidelines point towards the necessity of fairness and non-discrimination in the development, deployment and use of AI systems. For the EU-HLEG on AI (EC DG CNCT 2019), both the substantive and the procedural dimension of fairness are of importance. The substantive dimension pertains to the general commitment to fairness, which in various documents entails e.g. the equal and just sharing of benefits and costs (see e.g. APA 2022); the promotion of social justice and safeguarding of fairness and non-discrimination of any kind, making all reasonable efforts to minimise and avoid reinforcing or perpetuating discriminatory or biassed applications and outcomes (see UNESCO 2021a); and that the use of AI systems should never lead to people being deceived or unjustifiably impaired in their freedom of choice (EC DG CNCT 2019). The EU-HLEG on AI also points out that "fairness implies that AI practitioners should respect the principle of proportionality between means and ends, and consider carefully how to balance competing interests and objectives." (EC DG CNCT 2019: 12f)

The procedural dimension of fairness, on the other hand, ties into the steps laid out above, and entails the ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them. In order to do so, the entity accountable for the decision must be identifiable, and the decision-making processes should be explicable. As such, disclosure obligations in the development and application cycle and the attributability of responsibility and accountability are necessary requirements of a fair AI system (Heesen et al. 2020). It also means that human agency and oversight in the development and deployment of AI systems are essential components of fair AI systems.

ENSURING PRIVACY

The privacy standards as defined by the GDPR remain an important guide also in the context of GenAI systems. It lays out the need to safeguard individual rights regarding the use of personal data, and to provide easily accessible and understandable information about the data collection practices and the processing practices employed. In practice, this has a number of implications:

Since most GenAI applications are cloud-based, data must be uploaded to the cloud-infrastructure of the service provider in order to use these systems. Through this, a third party is given direct access to this data. By default, most cloud applications use uploaded data to further train their AI systems. Exploitable information includes e.g. text inputs, file uploads, and feedback, all of which might be triangulated with other information at a provider's disposal and used for purposes other than training the system. In addition, providers usually collect and store personal information on the user, which can also be linked to other data sources.

Not all data owners/sources permit AI-supported processing by use of an external AI system. Broadly speaking, published information such as texts on websites, publications, open data sets, patents, and codes can be used as input. However, it is important to verify whether the owner of the data allows or prohibits the input of these data, e.g. according to copyright or licensing agreements.

For personal data, informed consent for this type of processing must be sought. When planning to employ external AI systems e.g. for processing recordings or transcripts, informed consent sheets must reflect how and for what purposes these data are being processed and stored, and what measures will be taken to safeguard personal and sensitive data. Similarly, third party data such as non-published monitoring data of funding programmes provided by clients and project proposals submitted to funding agencies may only be used as an input if informed consent has been given. In general, confidential information, trade secrets, and data covered by non-disclosure-agreements may not be used as an input in external AI systems.

The same might not be true for local instances of an AI system with no connectivity to an external server or third-party service provider, where data security can be ensured. Still, these questions need to be clarified before employing such services, and relevant rules and regulations regarding a specific data handling task should be reviewed in advance. Likewise, any AI system's Terms of Service and Privacy Policy must be reviewed before inputting data, while organisational guidelines and the wishes of contractual parties/collaborators must be taken into account when deciding whether and under which conditions an AI system may be used.

As such, privacy necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols, as well as the capability to process data in a manner that protects privacy throughout the entire AI lifecycle. This is also true when employing AI in the implementation of evaluations, as it is crucial to ensure a proper handling of data and user privacy. In any case, GDPR-compliance must be ensured along every step of the process.

A NOTE ON INTELLECTUAL PROPERTY RIGHTS (IPR)

GenAI impacts innovation and creation and thus raises various questions regarding intellectual property (IP), including how IPR might intersect with AI models, their training data, their inputs and their outputs. The legal situation regarding the intellectual property rights of AI-generated outputs is currently very much contested and remains unsolved. The Terms of Use of a GenAI system may provide information where and how its outputs can be used. If the output of an AI system is used without significant changes, the origin of the outputs might need to be disclosed.

According to the World Intellectual Property Organization (WIPO 2024), incorporating AI systems into the innovation process can complement human creativity in generating ideas and solutions, while humans remain responsible for defining problems, setting goals, and determining how AI-derived insights are applied. However, the integration of generative AI in this context may complicate the licensing or patenting of new solutions. Copyright law, traditionally centred on human creators, generally restricts ownership to humans in many countries. Yet, this concept is being contested in some jurisdictions. For instance, early US rulings assert that AI-generated content is not eligible for copyright unless there is clear evidence of human creative contribution (Brittain 2023). In contrast, the Beijing Internet Court has recognised images generated by Stable Diffusion as original works, citing the human input through prompts as sufficient for copyright eligibility (Lai et al. 2024).

IPR is also not synonymous with Austrian copyright or "Urheberrecht", leading to further potential confusion. An added layer of complexity comes from the numerous ambiguities regarding human involvement and intentions (in terms of authorship) and the training data the AI system could have used in generating the output, often scraped from the internet without the consent of the original author(s) (El Atillah 2023). Finally, in the context of RTI (policy) evaluation, all evaluation results usually belong to the client, which makes the use of AI systems and ownership of results even more tricky.

CURRENT PRACTICES IN THE USE OF AI IN RESEARCH AND EVALUATION

AI will most likely have a considerable impact on the routine work of researchers and other experts engaged in the field of RTI evaluation. Overall, the percentage of researchers reporting extensive use of AI in their research increased from 12% in 2020 to 16% in 2021 across all fields – even before the wide release of ChatGPT-4 in March 2023. Bibliometric analysis shows consistent increase in the share of research papers mentioning AI or machine-learning terms across all fields over the past decade, reaching around 8% in total (van Noorden & Perkel 2023). A foresight study of the European Research Council (ERCEA 2023), conducted among ERC grantees, explored the concrete uses of AI in their scientific practice. The study shows that the use cases for AI systems are far-reaching, although they are still early in the process of being explored. Overall, respondents were positively inclined towards the opportunity for GenAI to handle repetitive or labour-intensive tasks (85%) and reduce language barriers (75%), although 93% of respondents found the implementation of ethical guidelines for AI a likely requirement. This tracks with serious concerns expressed regarding a lack of transparency and replicability (71% of respondents), but also the intrusiveness, manipulation and discrimination of AI systems (79% of respondents). AI systems also seem to be seen mostly as assistant technologies, although on average only about 54% of respondents were sceptical about the possibility of AI-based scientific publication and peer-review, while 50% saw AI affecting research integrity. Future perspectives on the use of AI are mostly collaborative and range from scientific discovery purposes, brainstorming on scientific ideas with increasingly responsive AI companions, to generating new scientific hypotheses. Only a limited number of researchers expect fully autonomous research processes, in which AI also takes actions such as selecting the best resources in the lab or accessing additional data if needed.

In the public policy arena, Tangi et al. (2022) report that policymakers recognise that the integration of AI in the public services could provide large benefits and public value to citizens (reduced administrative burden, personalised services, etc.), but they are also accompanied by some serious societal, ethical, regulatory, and technological challenges to be addressed. With respect to analytical purposes, potential use cases for analysis, monitoring and regulatory research, include 1) the process of inspecting, transforming, and modelling information into actionable knowledge (e.g. dashboard to support decision-making), 2) monitoring of policy implementation, and 3) processes for management of resources based on prediction models, to support planning. Furthermore, the study highlights potentials to increase public services and engagement management.

This makes apparent a number of benefits and challenges associated with the use of AI systems for evaluation purposes. In terms of benefits, mentioned dimensions include better cleaned data, possibility to analyse large quantities of data, more fluid processes through AI automation, real-time monitoring in rapidly changing contexts, removal of language barriers widening the access of STI research and policy with non-native speakers, better articulation of thoughts and enhanced communication, creative input of the AI, improvement of prompting, coding, programming skills, and most prominently, efficiency gains. (Ferretti 2023) Potential challenges, however, may reduce any such efficiency gains, for instance if there is a high dependence on a system and it fails to perform either in terms of content or technical aspects. Other challenges in the field of AI include ethical issues such as biases embedded in algorithms as well as their traceability, data security on servers beyond EU jurisdiction or data protection of personal information, as well as authorship and copyright concerns related to data used to train Large Language Models (LLMs). Additionally, incorporating AI into STI policy evaluation may present challenges due to its limited contextual understanding and hallucinations, potential reinforcement of the digital divide, and associated infrastructure and expertise costs. (Odumbe 2023)

AI USE WITHIN THE FTEVAL COMMUNITY

The survey developed by the working group for the fteval community led to a total of 20 responses, covering 14 of the 26 fteval member institutions. Within this group of respondents, funding agencies, research performing organisations and evaluators are relatively well represented, while ministries are covered the least. Still, these responses give important indications on attitudes and practices regarding the use of AI across relevant RTI organisations in Austria as of spring 2024. For instance, only 2 organisations indicated that no AI systems were in use at the time, neither at individual nor at organisation level. At the same time, only 4 respondents indicated that there were standard AI systems in use within their organisation – all of which were research performing organisations. 40% of respondents (8 out of 20) indicated that their institutions supplied them with certain AI systems for experimental use (representing 7 fteval members, including a ministry), while 75% of respondents (15 out of 20) feel the use of AI is mainly pushed by individuals. This seems to indicate that interest in AI was considerable, while use cases still needed to be established – according to respondents, no organisation used AI for decision-making processes. Only 3 of the responding organisations had no guidelines in place for the use of AI, with a fourth having not yet put them into writing. Of the remaining organisations, half had guidelines in development and half had them already in place.

Since adoption of AI was still rather early and there was at the time no clarity on relevant laws and regulations, any institutional guidelines must be seen as preliminary and could not cover all eventualities. While no responding organisation indicated a systematic use of AI systems within their institution, individual use of GenAI systems seemed widespread and hard to govern. The responses indicated that within the various organisations, some staff were already using AI systems for a lot of tasks, while others did not consciously use them at all. Complexities arise since AI systems pop up in different places unprompted, being integrated in ever more sophistication in search engines (such as Google and Bing), word processing software (as in Word assistant technologies), online meeting services (such as the Zoom AI Companion), graphic design and layout software (such as in the Adobe Suite), and operating systems (such as Microsoft Copilot). Nevertheless, the survey responses gave insight into typical scenarios in which AI systems are used or might be used in the future, broadly grouped by the authors into the following categories:

- **Text generation**, for instance for job vacancies, emails, programming/ coding, terms of reference (ToR), but also the development of survey items / questionnaires.

- **Proofreading and quality check**, including for accessibility/ plain language, coding, but also e.g. for the review of the fulfilment of all ToR or potentially for controlling on the use of AI for text and image generation.

- **Text and data processing**, including for the extraction, analysis and cleaning of data, summarising and synthesising, and for categorisation and text-mining.

- **Transcription** of audio, both live and recorded.

- **Translation of text**, which might also support simultaneous interviewing in different languages.
- **Knowledge gathering**, including desk research of relevant literature but also brainstorming, as an initial information tool to gain an overview, and for linking and mapping of thematic areas within a field.
- **Image generation**, including the creation of copyright-free images and document templates, visualisation of data and results, but also potentially to detect image manipulation.
- **Automatisation** especially of repetitive, monotonous, easy to control tasks, but potentially in the future also of processes such as peer-review.
- **Generalised support**, e.g. to enlarge participative formats, to serve as search tools in existing file systems, or for formatting reports.

Systems for language checking and translation – including DeepL – seem to be in regular use across institutions, while transcription tools were also mentioned a lot. More involved applications of GenAI, e.g. for summarising, structuring, mining, analysing or reworking texts, supporting desk research, developing codes for analysis, and phrasing of emails, were mentioned by most respondents as regular occurrences.

At the same time, a strong need for more guidance was expressed by most respondents. While the existing guidelines are mostly described as rudimentary – and dependent on future regulation – responsibility for exploring the use of AI systems, as well as for the downstream effects of said use, seemed at the time to be mostly passed onto the individual users. In this context, a provision of training was named as essential, especially on topics such as privacy and (GDPR) compliance, AI ethics, technical implementation, transparency, but also new developments in the field of AI and how they might be effectively employed. In this context, clear guidance is required both from ministries and employers that also outline responsibilities within the institution and beyond.

While a lot of existing institutional guidelines ask for AI generated outputs to be marked as such, there is no security on the extent to which AI is truly in use, and no standardisation in how such contents can and should be marked.

AI APPLICATIONS IN THE EVALUATION CONTEXT

The following section outlines some potential use cases for the application of AI systems for RTI evaluation by exploring synergies across the evaluation phases and discussing the feasibility of project and proposal evaluation, before proposing a self-assessment list to support the consideration process when adopting proprietary AI systems.

EXPLORING SYNERGIES AT THE INTERSECTION OF PROPRIETARY AI SYSTEMS AND EVALUATION PHASES

Experts working in the area of RTI (policy) evaluation perform a multitude of tasks in different organisational contexts. As researchers and evaluators, they are responsible for analysing RTI policies and programmes and their framework conditions within international, national, regional or sectoral innovation systems. As programme experts in national ministries, they are responsible for setting up R&I policies and instruments, delineating evaluation plans, and launching and overseeing R&I policy evaluation processes. Within funding agencies, RTI experts are responsible to design and execute transparent project selection mechanisms and ensure appropriate monitoring systems and portfolio analyses to guarantee efficient and effective funding, prevent fraud, and lay the foundations for impact measurement.

According to the 2019 fteval Standards (Kohlweg 2019), an evaluation typically comprises a preparation and planning phase, an implementation phase, and a management response phase. The implementation phase is further divided, with the term ‘inception phase’ commonly used when refining the evaluation design and reviewing its scope, purpose, methodology, and other aspects. The data collection phase, then, involves gathering relevant information to establish a basis for analysis and assessment. The data analysis and triangulation phase entail examining data from various sources, using diverse analytical techniques, and cross-referencing findings to improve the validity and reliability of conclusions. In the reporting phase, the findings, conclusions, and recommendations derived from data analysis are synthesised and communicated in a comprehensive and accessible manner to stakeholders.

To implement their tasks, experts from the evaluation community draw upon quantitative and qualitative research methods and tools to design policy interventions, set up monitoring and evaluation frameworks, implement monitoring and evaluation actions, and facilitate learning processes through engagement with diverse stakeholders. All these tasks and skills can be supported by AI to various degrees. According to Canadian evaluator and researcher Steve Jacobs, current literature suggests that AI could benefit every stage of the evaluation process (Jacobs 2023). In 2023, a group of nine evaluation experts analysed the evaluation activities and stages that could be affected by the use of artificial intelligence. On average, the study found that two thirds of evaluation activities can be impacted by artificial intelligence. (Head et al. 2023) The current level of support that AI can provide for evaluations is limited to narrow AI, which is task-specific. Some of these translate into applications which can be used at different stages in the implementation of RTI evaluations: 1) to systematically explore large or growing data sources (such as archives or document repositories) that are prohibitively time-consuming to process for humans; 2) to continuously improve assessments with new data being introduced (new categories and implementation challenges can be added, updated, etc. as the body of data to be assessed changes); and 3) to perform quality control and check for accuracy of methods employed. (Raimondo et al. 2023) Still, it is important to point out that at this stage, quality control remains a human task, and the AI user remains the responsible entity for ensuring the veracity and validity of any AI-generated outputs employed in their work. This is also an explicit requirement for the inclusion of AI applications in evaluative work for many institutions, including research funding organisations. The table below selectively maps use cases of AI capabilities and applications with evaluation phases and task examples. Since AI capabilities are constantly evolving, these must be seen as a snapshot.

AI capability	Application	Evaluation phase	Evaluation task examples
Natural Language Processing	Text editing, automated text generation	Inception, data collection & analysis, reporting & communication	Written deliverable (data collection tools, report, etc.), correspondence with evaluation stakeholders, coding, excel formulas, etc.

Natural Language Processing	Summarisation and synthesis	Data collection & analysis, reporting & communication	Tailoring evaluation results to cater to different audiences, synthesis of policies or previous studies for a review or a context section, etc.
Natural Language Processing and Machine Learning	Speech to speech, text to speech, text to text, speech to text	Data collection & analysis, reporting & communication	Automated dictation of interviews, focus groups, observations, evaluation meetings, live interpretation with end beneficiaries when no common language is spoken, translation of written deliverables
Natural Language Processing and Machine Learning	Document processing, data extraction	Data collection & analysis	Meta analyses, etc.

Natural Language Processing and Machine Learning	Chatbots, knowledge sharing	Inception, data collection & analysis	Refining, brainstorming on the best methodological approach, peer opinion, research companionship, collective intelligence gathering, desk research, design of interaction formats (training of evaluators, surveyors) sensitive to end beneficiaries. Social experiments such as through the simulation of social interactions, substitute or testing ground for human participants. (Grossman 2023)
Machine Learning	Predictive analytics	Data collection & analysis	Foresight and risk assessments based on historical data, etc.
Machine Learning	Anomaly detection	Data collection & analysis	Analysis of large sets of data such as financial transactions, outliers in results data, etc.

Table 1: Correspondence table of AI applications and evaluation tasks examples

PROJECT AND PROPOSAL EVALUATION

While the use of AI systems for the evaluation of projects and proposals can be seen as analogous to the higher-level evaluation outlined in the previous section, there are as of yet no established systems employed on a broader basis to support evaluation efforts. However, potential areas of application have been pointed out specifically to detect biases, inconsistencies and discrepancies in an evaluation process (Divasón et al. 2023a, 2023b), while also enhancing the speed and potentially overall scope of an evaluation, allowing for new insights to be drawn (Raimondo et al. 2023).

As the exploration of the potential use cases of AI for systematic employment in project and proposal evaluation are still at an early stage, only a few further considerations can be sketched out at this point:

- Legal and regulatory frameworks are still very much in the making. This needs to be taken into account when setting up new processes, in that they might need to be adjusted after the fact to fit new regulation.
- The costs of setting up new AI systems may be considerable, especially if they are to adhere to existing and developing regulation, including privacy standards and standards for scientific rigour and integrity.
- Big tech companies are thus in a favourable position, being able to invest large amounts of resources to the development of new technologies, while regulation is always necessarily lagging behind.
- The downstream effects of employing AI systems at a larger scale, especially to support decision-making processes, are currently hard to foresee. This might pertain to the development of inherently biassed systems, impacts on labour markets and work practices, and the overall quality of outputs in closed-loop systems.
- As such, it is important to proceed with care and commit to AI systems only if sufficient data is available on the consequences of such an institutional change.

PROPOSING A SELF-ASSESSMENT LIST TO CONSIDER PROPRIETARY AI SYSTEM ADOPTION

A semi-systematic evaluation of 22 guidelines (Hagendorff 2020) highlighted significant shortcomings in the field of AI ethics, noting that ethical guidelines often lack enforcement mechanisms, leading to no real consequences for deviations. At the time, ethics in AI was frequently viewed as a marketing tool rather than a core aspect of technology development, with guidelines having little impact on software developers' decision-making. There was a general perception of AI ethics as an optional addition, and a distributed responsibility model diluted accountability. Economic incentives often outweighed ethical commitments, suggesting a misalignment between AI development purposes and societal values. Despite these challenges, there were efforts to address specific ethical concerns in AI through technical solutions, such as enhancing privacy, anti-discrimination measures, and explainability. However, many ethical issues remain under-addressed or ignored in guidelines, spanning a wide range of topics from the potential for malevolent AI to the social and ecological impacts of technology. The study suggested the need for more comprehensive and effectively enforced ethical guidelines in AI development, which remain relevant today as the range of stakeholders using AI systems has expanded to include the wider public. Following this, in 2020 the European Commission published a guide on ethical principles for designers and developers of AI systems, data scientists, procurement officers or specialists, front-end staff working with AI systems, legal and compliance officers, and managers, against which they can self-assess the trustworthiness of AI systems (EC DG CNCT 2020).

Beyond ethical considerations, more concrete operational steps should be envisaged as part of a process using an AI system or purchasing AI services. These are outlined for consideration in the figure below:

- **Plan thoroughlly by reflecting on evaluation phases to be supported, tasks to be performed by the AI system, and flexibility needed.**
- **Assess the availability, quantity, and quality of the data to be analysed to ensure effificency.**
- **Review existing data systems and software with which the AI should be compatible and check for scalability in case of increased workload.**
- **Estimate the budget comprehensively, including licenses, updates and maintenance, training needs for non-specialists, and vendor support options.**
- **Examine compliance with ethical and legal frameworks, security and privacy features.**
- **Consider ease of use by testing the tool and comparing effecteviness with alternatives using free trials. Invest time in reviews and case studies to avoid pitfalls.**

Figure 2: Self-assessment considerations for adopting proprietary AI systems

OUTLOOK

The technical, legal and ethical competencies needed in the evaluation community to utilise AI systems responsibly will increase. To explore the opportunities and the ways to employ AI systems responsibly for better evaluations, AI skills training and open online resources should be built with a focus on technical, legal and ethical questions. As outlined in the section on AI use in the fteval community, there is a strong need for more guidance and training to support the responsible and effective application of AI systems for evaluation purposes. Key topics for which training is needed broadly cover best practices on the development, deployment and application of AI systems, again focussing on technical, legal and ethical aspects such systems imply. In particular, questions on privacy and (GDPR) compliance, transparency, as well as quality assurance, ethics, and bias loom large. Here, the fteval platform might provide support to member organisations by inviting external experts or continuing exchange meetings, allowing the community to keep track of this quickly changing field.

REFERENCES AND FURTHER READING

ACRAI – Österreichischer Rat für Robotik und Künstliche Intelligenz. (2018). Die Zukunft Österreichs mit Robotik und Künstlicher Intelligenz positiv gestalten. White Paper des Österreichischen Rats für Robotik und Künstliche Intelligenz. https://www.bmk.gv.at/dam/jcr:f2f7a973-8aa4-4be8-9a6b-0c7c44e73ce4/white_paper_robotikrat.pdf.

Ada Lovelace Institute. (2023). Regulating AI in the UK: Strengthening the UK's proposal for the benefit of people and society. https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/.

AI-Kompetenznetzwerk. (2023). Kompass für den dienstlichen Umgang mit generativer Künstlicher Intelligenz (KI). Magistratsdirektion der Stadt Wien. Geschäftsbereich Organisation und Sicherheit. Gruppe Organisation. Wien.

APA – Austria Presse Agentur. (2022). Leitlinie zum Umgang mit künstlicher Intelligenz. Wien.

Biegelbauer, P., Lackinger, C., Schlarb, S., Subak, E., Weinlinger, P. (2023). Leitfaden Digitale Verwaltung und Ethik. Praxisleitfaden für KI in der Verwaltung, Version 1.0. Bundesministerium für Kunst, Kultur, öffentlicher Dienst und Sport. Wien.

BMK. (2021a). Strategie der Bundesregierung für Künstliche Intelligenz. Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie. Wien.

BMK. (2021b). Strategie der Bundesregierung für Künstliche Intelligenz: Annex. Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie. Wien.

BMWE. (2020). Ethik und Künstliche Intelligenz. Was können technische Normen und Standards leisten? Whitepaper. DIN e.V. und DKE. Berlin.

Bockting, C. L., et al. (2023). Living guidelines for generative AI — why scientists must oversee its use. In: Nature Vol. 622.

Brittain, B. (2023). AI-generated art cannot receive copyrights, US court says. Reuters. https://www.reuters.com/legal/ai-generated-art-cannot-receive-copyrights-us-court-says-2023-08-21/ (last accessed 24.09.2024).

BVDW. (2019). Acht Leitlinien für künstliche Intelligenz. Leitlinien des BVDW. Bundesverband Digitale Wirtschaft. Berlin.

de Witt, C., Rampelt, F., Pinkwart, N. (Hrsg.). (2020). Künstliche Intelligenz in der Hochschulbildung – Whitepaper. Berlin: KI-Campus. https://doi.org/10.5281/zenodo.4063722.

Der Bundesrat. (2020). Leitlinien «Künstliche Intelligenz» für den Bund: Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung. Schweizerische Eidgenossenschaft.

Der Bundesrat. (2022). Monitoring der Leitlinien «Künstliche Intelligenz» für den Bund: Evaluation der Anwendung und Aktualität der Leitlinien. BAKOM-Bericht. Schweizerische Eidgenossenschaft.

DFG. (2023). Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG.

Divasón, J., Martínez-de-Pisón, F.J., Romero, A., Sáenz-de-Cabezón, E. (2023a). Artificial Intelligence Models for Assessing the Evaluation Process of Complex Student Projects. In: IEEE Transactions on Learning Technologies 16(5). 694-707. https://doi.org/10.1109/TLT.2023.3246589.

Divasón, J., Martínez-de-Pisón, F.J., Romero, A., Sáenz-de-Cabezón, E. (2023b). Robustness Analysis of a Methodology to Detect Biases, Inconsistencies and Discrepancies in the Evaluation Process. In: García Bringas, P., et al. International Joint Conference 16th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2023) 14th International Conference on EUropean Transnational Education (ICEUTE 2023). CISIS ICEUTE 2023. Lecture Notes in Networks and Systems, vol 748. Springer, Cham. https://doi.org/10.1007/978-3-031-42519-6_31.

Europäische Kommission, Generaldirektion Kommunikationsnetze, Inhalte und Technologien. (2019). Ethik-Leitlinien für eine vertrauenswürdige KI. Amt für Veröffentlichungen der Europäischen Union. https://doi.org/10.2759/22710.

Europäische Kommission. (2022). Ethische Leitlinien für Lehrkräfte über die Nutzung von KI und Daten für Lehr- und Lernzwecke. Amt für Veröffentlichungen der Europäischen Union. https://doi.org/10.2766/494.

EC – European Commission. (2020). White Paper on Artificial Intelligence—A European approach to excellence and trust. COM(2020).

EC DG CNCT – European Commission, Directorate-General for Communications Networks, Content and Technology. (2019). Ethics guidelines for trustworthy AI. Publications Office. https://data.europa.eu/doi/10.2759/346720.

EC DG CNCT – European Commission, Directorate-General for Communications Networks, Content and Technology. (2020). The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment. Publications Office. https://data.europa.eu/doi/10.2759/002360.

EC DG RTD – European Commission, Directorate-General for Research and Innovation, Arranz, D., Bianchini, S., Di Girolamo, V. et al. (2023). Trends in the use of AI in science – A bibliometric analysis. Publications Office of the European Union. https://data.europa.eu/doi/10.2777/418191.

El Atillah, I. (2023, July 10). Copyright challenges in the age of AI: Who owns AI-generated content? Euronews. https://www.euronews.com/next/2023/07/10/copyright-challenges-in-the-age-of-ai-who-owns-ai-generated-content (last accessed 24.09.2024).

ERCEA – European Commission, European Research Council Executive Agency. (2023). Use and impact of artificial intelligence in the scientific process – Foresight. Publications Office of the European Union. https://data.europa.eu/doi/10.2828/10694.

Ferretti S. (2023). Hacking by the prompt: Innovative ways to utilize ChatGPT for evaluators. In: New Directions for Evaluation 178-179. 73-84. https://doi.org/10.1002/ev.20557.

Feuerriegel, S., Hartmann, J., Janiesch, C., Zschesch, P. (2024). Generative AI. In: Business & Information Systems Engineering 66. 111-126. https://doi.org/10.1007/s12599-023-00834-7.

Google. (2022). AI Principles Progress Update.

Grossmann, I. (2023, July 3). Beyond the hype: How AI could change the game for social science research. The Conversation. https://theconversation.com/beyond-the-hype-how-ai-could-change-the-game-for-social-science-research-208086 (last accessed 24.09.2024).

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. In: Minds and Machines 30. 99-120. https://doi.org/10.1007/s11023-020-09517-8.

Head, C. B., Jasper, P., McConnachie, M., Raftree, L., Higdon, G. (2023). Large language model applications for evaluation: Opportunities and ethical implications. In: New Directions for Evaluation 178-179. 33-46. https://doi.org/10.1002/ev.20556.

Heesen, J. et al. (Hrsg.). (2020). Ethik-Briefing. Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen – Whitepaper aus der Plattform Lernende Systeme. München.

Heidrich, J. (2023). Gut ausgerichtet: Richtlinien für die Nutzung von KI im Unternehmen schaffen. In: c't 2023, Heft 19.

Heller-Schuh, B., Kasztler, A., Leitner, K.-H. (2019). Künstliche Intelligenz als thematische Herausforderung für österreichische Universitäten. AIT-IP-Report Vol. 21.

Hense, J., Rädiker, S. (2023). ChatGPT & Co. – Anwendungsszenarien künstlicher Intelligenz in der Evaluation. Foliensatz. https://www.degeval.org/fileadmin/content/Arbeitskreise/AK_Methoden/KI/JT2023_AK-Session_A4_KI_Doku.pdf.

Kohlweg, K. (2019). Evaluation Standards for Research, Technology and Innovation Policy. Technischer Bericht. fteval - Österreichische Plattform für Forschungs- und Technologiepolitikevaluierung. Wien. https://doi.org/10.22163/fteval.2019.344.

Jacobs S. (2023, December 12). Webinaire sur l'intelligence artificielle et le futur de l'évaluation. Société canadienne d'évaluation (SCÉ) & Perfeval.

Lai, S. Lim, D. Shi, L., Tay, J. (2024, March). Legal implications – Beijing Internet Court grants copyright protection to AI-generated artwork. National University of Singapore – Centre for Technology, Robotics, Artificial Intelligence & the Law. https://law.nus.edu.sg/trail/legal-implications-beijing-internetcourt-copyright/ (last accessed 2024.09.24.).

MacLaren, I., O'Brien, G., et al. (2023). Generative Artificial Intelligence: Guidelines for Educators. National Academic Integrity Network. Quality & Qualifications Ireland.

Nancholas, B. (2023, October 12). Natural language processing: An explainer. University of Wolverhampton. https://online.wlv.ac.uk/natural-language-processing-an-explainer/ (last accessed 2024.09.24.).

NSAI. (2023). AI Standards & Assurance Roadmap. National Standards Authority of Ireland. Dublin.

Odumbe, K. O. (2023, December 11). Integrating AI into Monitoring and Evaluation: A Pathway for Enhanced Efficiency in Development Work. LinkedIn. https://www.linkedin.com/pulse/integrating-ai-monitoring-evaluation-pathway-enhanced-ken-odak-odumbe-u3pef/ (last accessed 2024.09.24.).

OECD. (2022). Artificial Intelligence in Science. Challenges, Opportunities and the Future of Research. OECD Publishing. Paris. https://doi.org/10.1787/a8d820bd-en.

OECD. (2024 [2019]). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.

OECD. (2024). Transparency and explainability (Principle 1.3). OECD AI Principles. https://oecd.ai/en/dashboards/ai-principles/P7 (last accessed 24.09.2024).

Ostendorf, A., Peters, M. (2003). Handreichung für Studierende an der Fakultät für Betriebswirtschaft zum Einsatz von KI-Tools im Studium. Leopold-Franzens-Universität Innsbruck.

Pant, A. (2019, January 7). Introduction to Machine Learning for Beginners. Towards Data Science. Medium. https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08 (last accessed 2024.09.24.).

Pinaya, W. H. L., Graham, M. S., Kerfoot, E., Tudosiu, P.-D., Dafflon, J., Fernandez, V., Sanchez, P., Wolleb, J., da Costa, P. F., Patel, A., Chung, H., Zhao, C., Peng, W., Liu, Z., Mei, X., Lucena, O., Ye, J. C., Tsaftaris, S. A., Dogra, P., Feng, A., Modat, M., Nachev, P., Ourselin, S., Cardoso, M. J. (2023). Generative AI for Medical Imaging: extending the MONAI Framework. In: Electrical Engineering and Systems Science. https://doi.org/10.48550/arXiv.2307.15208.

Poretschkin, M. et al. (2021). Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz. KI-Prüfkatalog. Fraunhofer IAIS. Sankt Augustin.

Projektgruppe „Künstliche Intelligenz" des VDMA Bayern. (2020). Leitfaden Künstliche Intelligenz – Potenziale und Umsetzungen im Mittelstand. VDMA Bayern. München.

Prunkl, C. (2024). Human Autonomy at Risk? An Analysis of the Challenges from AI. In: Minds & Machines 34, 26. https://doi.org/10.1007/s11023-024-09665-1.

Raimondo, E., Anuj, H., Ziulu, V. (2023, August 16). Setting up Experiments to Test GPT for Evaluation. Independent Evaluation Group – World Bank Group. https://ieg.worldbankgroup.org/blog/setting-experiments-test-gpt-evaluation (last accessed 24.09.2024).

Regona, M., Yigitcanlar, T., Xia, B., & Li, R. Y. M. (2022). Opportunities and Adoption Challenges of AI in the Construction Industry: A PRISMA Review. In: Journal of Open Innovation: Technology, Market, and Complexity 8(1). 45. https://doi.org/10.3390/joitmc8010045.

Russell Group. (2023). Russell Group principles on the use of generative AI tools in education.

Segreto, T. (2014). Knowledge-Based System. In: Laperrière, L., Reinhart, G. (eds) CIRP Encyclopedia of Production Engineering. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-20617-7_6557.

Tangi L., van Noordt C., Combetto M., Gattwinkel D., Pignatelli F. (2022). AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector. EUR 31088 EN. Publications Office of the European Union. Luxembourg. https://doi.org/10.2760/39336.

U.S. Department of Education, Office of Educational Technology. (2023). Artificial Intelligence and the Future of Teaching and Learning. Insights and Recommendations. Washington, DC.

UNESCO. (2021a). Recommendation on the ethics of artificial intelligence. UNESCO France.

UNESCO. (2021b). Ethical impact assessment: a tool of the Recommendation on the ethics of artificial intelligence. UNESCO France. https://doi.org/10.54678/YTSA7796.

Van Noorden, R., & Perkel, J. M. (2023). AI and science: What 1,600 researchers think. In: Nature 621(7980). 672–675. https://doi.org/10.1038/d41586-023-02980-0.

WIPO. (2024). WIPO Conversation IP and Frontier Technologies: Generative AI. https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_conv_ge_2_23/wipo_ip_conv_ge_2_23_summary.pdf.

Wischmann, S. et al. (2022). Leitfaden für das Qualitätsmanagement bei der Entwicklung von KI-Produkten und -Services. Begleitforschung des Technologieprogramm KI-Innovationswettbewerb des Bundesministeriums für Wirtschaft und Klimaschutz (BMWK). Institut für Innovation und Technik (iit) in der VDI/VDE Innovation + Technik GmbH. Berlin.

AUTHORS

STEFANIE SCHUERZ
ZSI - Zentrum für Soziale Innovation GmbH
Linke Wienzeile 246
1150 Vienna
ORCID 0000-0002-3816-1294

CHARLOTTE D'ELLOY
Technopolis Group
Rudolfsplatz 12/11
1010 Vienna

MICHAEL DINGES
AIT Austrian Institute of Technology GmbH
Giefinggasse 4
1210 Vienna
ORCID 0000-0003-0433-4318

ANNEX I: QUESTIONNAIRE TO THE FTEVAL COMMUNITY

The following questionnaire was translated from the original German into English.

1. **Information on you and your organisation**
    - What is your name? (Open question)
    - In what organisation are you employed? (Open question)
    - What is your role within the organisation? (Open question)

2. **Relevance of AI in your organisation and your area of work**
   (Likert scale: 1. not relevant; 2. scarcely relevant; 3. neither/nor; 4. relevant; 5. very relevant)
    - How relevant is the topic of AI for your organisation at this time?
    - How relevant is the topic of AI for you in your area of work?
    - How relevant is the topic of "Guidelines for the use of AI" in your organisation?
    - How relevant are "Guidelines for the use of AI" in your area of work?

3. **In what stage of AI use is your organisation?**
    - AI tools are not used in our organisation – neither by individuals nor on an organisational level.
    - AI tools are mainly used by individuals for their own purposes.
    - Our organisation has provided certain AI tools for trial in specific work areas.
    - There are AI tools that are used by default in our organisation.
    - Other:

4. **Are there already guidelines for the use of AI in your organisation?**
    - Yes
    - No
    - Currently in development
    - I don't know

5. If yes or in development: What experiences are there in your organisation regarding "Guidelines for the use of AI"? (Open question)

6. Are AI tools used in your organisation or in your area of work for tasks that are part of decision-making processes?

- Yes
- No
- I don't know

7. If yes, which tasks? (Open question)

8. What experiences are there with the use of AI in your organisation or in your area of work?

9. What is a typical scenario in your area of work where you are using AI tools, or might do so in the future? (Open question)

10. Please list typical AI tools that you are using or that you know are being used in your organisation. What are these tools used for?

11. Which kind of support would you like to have to ensure a safe and reliable application of AI as an employee?

12. How do you handle receiving AI-generated outputs from external stakeholders (e.g. project partners, beneficiaries)?

13. What are the biggest challenges you or your organisation are currently facing with regard to the use of AI tools?

14. Any other comments?

ANNEX II: AI SYSTEMS

Name	Description	Evaluation step	Pricing model	Language support	Alternative	Review
consensus	Consensus is a search engine that uses language models to surface papers and synthesize insights from academic research papers. The current source material used in Consensus comes from the Semantic Scholar database, which includes over 200M papers across all domains of science. More papers are meant to be added. The dataset is updated on a monthly cadence. Consensus is not a chatbot, but uses the same technology to help make the research process more efficient.	- Proposal writing/ context - Assessment of the evaluation criteria "relevance of programme design"	Freemium	English only	Elicit, Research GPT, ScholarAI, Scisummary, HeyScience	https://www.youtube.com/watch?v=YIow-IQFS9rg
fireflies.ai	Fireflies.ai is an AI transcription tool that records and transcribes meetings across various web-conferencing platforms like Teams, Zoom and Google Meet, actively listening and noting key insights converting spoken words into text and generating summaries. Besides basic transcription, Fireflies.ai also analyzes meeting sentiments, categorizing them into positive, negative, and neutral segments for easier review and team sharing.	- Data collection: interviews/ focus groups - Project management: client meeting	Freemium	60+ languages, incl. German, Dutch, French, Spanish, Portuguese, Italian and three English accents: UK, Australian, and US.	Airgram.io, Tactiq, Whisper	https://www.notta.ai/en/blog/fireflies-ai-review
Curie	Curie is an AI-powered writing assistant designed specifically for academic papers. It provides intelligent suggestions, improves writing structure, enhances flow, and assists with citations to help produce high-quality academic content. Curie uses advanced artificial intelligence and natural language processing algorithms to assist you in editing and translating scholarly writing. Curie analyses the input provided by users and suggests edits, helping with tasks like drafting articles, polishing grant applications, or improving writing style. Curie has been specifically designed for research writing, trained on a specialized collection of manuscripts edited by professional subject editors.	Report drafting	Free trial for 14 days Premium $11.25 Custom price for a preferred group of people	English only translations from Chinese, Portuguese and Spanish	PaperPal, Grammarly, Quillbot, writefull X, compose.ai	-

Name	Description	Evaluation step	Pricing model	Language support	Alternative	Review
scite_	Scite_ is a citation index tool that takes advantage of recent advances in artificial intelligence to produce "Smart Citations." Smart Citations reveal how a scientific paper has been cited by providing the context of the citation and a classification system describing whether it provides supporting or contrasting evidence for the cited claim, or if it just mentions it.	Report drafting	Monthly €15,92, yearly €114,63. Additionally there are a few free applications (e.g. browser extension)	English (no information on other languages)	Zotero, Mendeley, Citavi, Jotbot	MIT Press
Browse AI	Browse AI is a web-based tool that allows you to extract and monitor data from any website without coding. With Browse AI, you can train a robot by clicking on the elements you want to extract from a website. The robot will simulate your actions and extract your desired data on your chosen schedule, giving you a live web data pipeline within minutes.	Data collection and monitoring	Freemium	English (no information on other languages)	Fivetran, hevo DATA, Apify, Dataddo, Bright data	Browse ai Erfahrung: Automatisieren mit der KI. Ist es gut? (seo-tech.de)
CHATPDF	ChatPDF simplifies reading and interacting with PDF documents by allowing users to ask real-time questions and receive context-specific answers based on the document's content. Utilizing natural language processing and deep learning, it enables a conversational interface with PDFs by analyzing the text within uploaded documents to understand and extract key concepts. ChatPDF provides detailed answers by matching questions to relevant information in the document, supporting follow-up questions for an interactive and efficient experience. This tool combines scanning, natural language understanding, and AI-driven response generation to enhance document comprehension and interaction.	Proposal writing, data collection	Freemium	ChatPDF accepts PDFs in any language and can chat in any language.	Elephas APP, PDFgear	ChatPDF Review (2024): Should You Try It? | Great Software

Name	Description	Evaluation step	Pricing model	Language support	Alternative	Review
	Quivr is designed as a personal AI database or "second brain," helping users store and effortlessly retrieve unstructured information through an AI knowledge base built on their data. Quivr is an exception in this list since it's open-source, and can be run locally, ensuring transparency in data storage and security. Powered by generative AI, Quivr automatically organizes and categorizes uploaded information, simplifying access and eliminating the need for manual sorting. This tool is ideal for those inundated with information daily, helping clear mental clutter for better focus. Quivr offers a demo, access to its GitHub for customization, and a Discord channel for community support, emphasizing its role in streamlining information management with AI.	Data collection, data analysis	Freemium	No information	Obsidian	https://medium.com/@manaaki.walker-tepania/quivr-your-personal-conversational-knowledge-base-harnessing-the-power-of-large-language-3785ec8e3f96
	Rows is the only spreadsheet with the capabilities of an AI Analyst and a native OpenAI integration. Both are available to everyone and free to use. What makes the AI features in Rows special is the combination of the spreadsheet interface with the AI capabilities. This means users can leverage AI to drag automations across ranges of cells, combine the AI actions with any other spreadsheet formula, and use cell references to create dynamic AI experiences. Addiationlly, the ability to use AI to summarize and help you make sense of data is special. Rows offers the only product that allows users to ask questions about a dataset using plain language.	Data collection	Freemium	English (no information on other languages)	Sheet AI	-

Name	Description	Evaluation step	Pricing model	Language support	Alternative	Review
formularizer	Formularizer is an AI-powered assistant that streamlines the creation and understanding of formulas, scripts, and Regex patterns for Excel, Google Sheets, Notion, and other platforms. It ensures user data security with the latest technologies, without storing any data on its servers. Leveraging OpenAI's advanced GPT-4 model, this tool simplifies turning text instructions into precise formulas rapidly and for free, supporting various operations and allowing for regular expression use.	Data analysis	Formularizer is free for 150 uses each month (5 uses per day), which is enough for most users. There also is a Premium version.	No information	Rows, GPTExcel, Excelformulabot	-
iThenticate 2.0	iThenticate 2.0 is the new plagiarism detection programme by turnitin. This programme is characterised by its high sensitivity and accuracy in detecting plagiarism and AI-generated text. iThenticate 2.0 is designed to help with quality assurance in science and in the production of content in high-stakes areas, such as government institutions. iThenticate 2.0 introduces a revitalized and contemporary interface, crafted with the research community in mind, and enabling effortless navigation from the initial onboarding stage through to the final similarity check.	Quality assurance	$100 starting costs	English, German, Spanish, Japanese and more	Grammarly Business, GPTZero	iThenticate Pricing, Alternatives & More 2024 | Capterra