AIIA 2.0 for KiM Explorer

11 november 2025

Blue questions are mandatory. Green questions are intended to provide additional information. They should be completed if they are relevant.

Part A: Assessment

1 System purpose and necessity

1.1 Purpose of the system

1.1.1 Provide a brief description of the intended purpose and intended result of the AI system (title, general description, definition of the problem, expected timeframe, location, target groups, the domain and operational process).

- Title: KiM Explorer.
- General description: An AI-based, web-based chatbot that uses retrieval-augmented generation (RAG) to make research publications from the Netherlands Institute for Transport Policy Analysis (KiM) searchable and interactive.
- Problem definition: As KiM's archive of publications grows, it becomes increasingly difficult for researchers, policymakers, and the public to find relevant knowledge using traditional keyword searches on titles and descriptions.
- Expected result: To allow users to search the full text of publications using natural language, find relevant documents semantically, and interact with the content of those documents through a conversational interface to get cited answers to their questions.
- Target groups: KiM researchers (currently), policymakers (limited pilot), and general public (potential future)
- Domain: Dutch transport policy analysis.

See https://gitlab.com/kennisinstituut-voor-mobiliteitsbeleid/kim-explorer

1.1.2 In which risk level in the AI Regulation does your AI system fall: unacceptable, high or minimal risk?

The KiM Explorer falls into the minimal risk category. It is a tool for information retrieval from publicly available documents and does not make automated decisions that have a legal or otherwise significant effect on individuals. Disclaimers are provided on required human validation for using the information elsewhere.

1.1.3 Where in the organisation (in which processes) is the AI system intended to be used?

The system is used by the KiM Netherlands Institute for Transport Policy Analysis to make its own public literature more accessible for internal and external research and policy analysis processes. It will be expanded to a limited pilot with a known, preselected list of users within specific policy departments of the Ministry of Infrastructure and Water Management (IenW → DGMo/DGLM), who will be able to use the application to support their decision-making, with a clear disclaimer that all information must be validated in the source text before being used or reproduced.

1.2 Intended solution

1.2.1 Provide a brief description of the intended AI system (technology, data and type of algorithm).

- Technology: The system uses a two-stage Retrieval-Augmented Generation (RAG) architecture. It leverages a Large Language Model (LLM) and vector search via the

OpenAI API. The front-end is built with Python and the NiceGUI framework.

- Data: The input data consists of the research publications published by KiM, stored in an OpenAI vector store. This includes PDFs and other documents, along with their metadata (title, authors, date, etc.).
- Algorithm:
  - Search stage: A user's query is rewritten by an LLM for optimization. A vector search then identifies relevant document chunks. These findings are summarized by the LLM and presented to the user.
  - Interaction stage: The user selects specific publications from the search results. The full text of these selected documents is then provided as context to an LLM to generate comprehensive, cited answers to the user's initial and follow-up questions.

See the project brief on https://gitlab.com/kennisinstituut-voor-mobiliteitsbeleid/kim-explorer/-/blob/main/docs/brief.md

1.2.2 Why was this form of AI chosen (e.g. generative AI, linear regression or neural network)?

The two-stage RAG architecture was deliberately chosen for two main reasons:

- Transparency & user control: It makes users aware of the exact scope of the information being used to generate answers by requiring them to actively select the source documents.
- Accuracy & thoroughness: Using the full text of selected documents as context, rather than just the most relevant chunks, leads to more accurate and comprehensive answers, which is crucial in a scientific context.

1.2.3 What alternatives were considered (e.g. no AI, less complex AI, different type of algorithm)?

This system complements the existing, AI-less approach of manual keyword search and contact with KiM colleagues. As time passes and the data library grows, this becomes increasingly inefficient. This is where an AI system could fill a gap. It would also make it easier to relate and compare different publications on the same topic.

1.3 Role within the organisation

1.3.1 Describe the division of tasks in setting up the AI system (such as the developer, commissioning client, project leader, IT management organisations and person with ultimate responsibility). If an external party is responsible for development: what contractual agreements are in place?

The system was developed by the KiM Netherlands Institute for Transport Policy Analysis internally. Git history provides who contributed what, when and why.
https://gitlab.com/kennisinstituut-voor-mobiliteitsbeleid/kim-explorer/-/commits/main

1.3.2 Who will be the user of the AI system, who are the end users who will work with the system and which parties involved will be impacted by the AI system?

- End users: KiM researchers (currently), policymakers (limited pilot), and the general public (possibly in the future).
- Impacted parties: The same groups are impacted by having improved access to KiM's research findings. Dutch travelers and the logistics sector benefit from a better underpinned transportation policy.

1.3.3 Which stakeholders, people and/or groups have been consulted in the development of the AI system?

KiM colleagues (mobility researchers) have been consulted, and ideas/best practices have been

exchanged with staff focusing on AI at other government knowledge institutes (RKIs) and the ministry.

1.3.4 What feedback has been collected from teams or groups representing different backgrounds and experiences? And how was this feedback followed up?

The initial proof-of-concept has been extensively tested within KiM. Based on the feedback received by colleagues, the functionality and system prompt have been improved, e.g. to ensure better response quality, limit unwanted behaviour, and clarify parts of the interface. Feedback was tracked and resolved on a public issue tracker.

1.4 Maintenance and administration

1.4.1 Describe the division of tasks for the administration and maintenance of the AI system (such as the developer, commissioning client, project leader, management organisations and party with ultimate responsibility). If an external party is responsible for developing the system: what contractual agreements are in place?

The system is an open-source project managed by KiM. Roles and tasks are divided as needed, and can be largely tracked in the GitLab issue tracker: https://gitlab.com/kennisinstituut-voor-mobiliteitsbeleid/kim-explorer/-/issues?sort=created_date&state=all Maintenance and administration is handled in-house by KiM, alongside ODC-Noord, the organisation where hosting is contracted.

1.4.2 How are new laws and regulations that may be introduced or updated during the lifetime of the AI system taken into account?

The KiM AI team has regular meetings discussing the state of affairs of AI at KiM, the ministry, the government and beyond – including updated regulations or frameworks affecting this application.

1.4.3 Has the expertise required to manage the AI system been documented?

Yes, the code structure, component breakdown, and file overview in Structure.md and the general README provide a clear picture of the system architecture. This implicitly documents the required expertise in Python, NiceGUI, and the OpenAI API.

1.4.4 How are changes in the context of the AI system taken into account?

The context is the library of KiM publications. This context is updated by extending the OpenAI vector store with new or revised documents. This is automated in CI, based on the documents on the KiM website. These documents only include publications that are produced and validated in-house by KiM.

2. Impact

2.1 Fundamental rights

2.1.1 What will the potential impact be on citizens' fundamental rights of using the AI system?

The potential impact on fundamental rights is minimal. The system operates on publicly available information and does not make decisions affecting individuals. The primary, low-level risk is the potential for the LLM to provide inaccurate information, which could impact the right to receive correct information. This risk is mitigated by the system's design, which provides direct links to the source documents for verification.

2.1.2 What legal basis underlies the use of the AI system and the intended decisions to be taken based on the AI system?

n/a

2.1.3 Which constitutional provisions may be applicable?

n/a

2.1.4 Which of these constitutional provisions may be breached in the event of improper implementation of the AI system?

n/a

2.2 Sustainability

2.2.1 What will be the environmental impact of introducing the AI system (development, installation and use), and how will this be measured?

The application itself is hosted at ODC-Noord, using a server of the same type as other applications by KiM. This server represents about 23% of current total hosting capacity contracted by KiM.
The RAG and LLM components of the application are provided by OpenAI. We plan to log the amount of queries and tokens used, although this will not be directly linked to environmental impact.

2.2.2 What measures have been taken to minimise the (negative) environmental impact of the AI system?

The data centre used by ODC-Noord uses 100% renewable electricity and aims to be entirely CO_2-neutral in its operations by 2030.
For the RAG and LLM components, we have less control about energy usage or other forms of environmental impact. We do opt for the smallest possible models while still providing acceptable performance. Smaller models are cheaper and consume less energy than larger models.

2.3 Other effects

2.3.1 How does the AI system contribute to the organisation's mission?

It directly supports KiM's vision that its publications should be "findable, discoverable and easily accessible." It transforms the static publication archive into a dynamic, interactive knowledge base, thereby enhancing the dissemination and utility of KiM's research.

2.3.2 In addition to the questions above, are there any other relevant effects (positive, negative, risks, for specific target groups, at different levels, broad prosperity) of the AI system that need to be taken into consideration?

- Positive effects: Increases the efficiency of research for policymakers and academics, leading to better-informed policy decisions and literature research. Makes complex research more accessible to the public.
- Negative effects/risks: Risk of user over-reliance on AI-generated summaries without consulting the original, nuanced research. Potential for LLM hallucinations or misinterpretations presenting inaccurate information, despite design mitigations. A potential for deskilling in traditional literature search methods.

3. Assessing whether or not to use the AI system

3.1 Is the impact in proportion to the intended goals and are there other less radical ways of achieving these goals? In other words: is it proportional and in line with subsidiarity to deploy the system to achieve the stated goals?

Yes, the use of the system is proportional. The goal is to overcome the limitations of keyword search in a large, growing body of literature. A semantic, natural-language-based AI system is a direct and effective solution. The chosen two-stage architecture, which puts the user in control of source selection, is a less radical approach than a fully automated system, ensuring the intervention is appropriate for the goal of providing transparent and accurate information retrieval.

3.2 Are there additional measures (e.g. as part of processes) that you could take to use the system responsibly?

The system already incorporates key responsible AI measures by design: • Human-in-the-loop: Users actively select the source documents. • Transparency & citations: The system clearly indicates its sources and provides links to them. • Disclaimers: The UI is designed to include warnings about AI-generated content.

Part B: Implementation and use of AI system

4. Technical robustness

4.1 Bias

4.1.1 How will potentially undesirable bias, such as bias in the input, bias in the model and bias in the output of the AI system be taken into account?

Currently: • Input bias: The primary source of bias is the content of the KiM publications themselves. The system does not alter this, only makes it accessible. • Model bias: The system (currently) uses a third-party LLM (from OpenAI) and is subject to its inherent biases. • Output bias: The system's architecture, which grounds responses in user-selected source texts, is a significant mitigation against the model introducing external biases or fabricating information. Future: We plan to implement an objective benchmark to function as ground truth (working name: KiM-bench). This allows us to evaluate and monitor the completeness, accuracy and neutrality of answers.

4.1.2 Is the input (data) relevant and representative, taking account of the intended purpose (question 1 of 1.1) of the AI system?

Yes: we aim to make all data published by KiM (and currently only that data) accessible; and all this data has indeed been included in the dataset.

4.1.3 In random sampling, have any subpopulations been protected if necessary?

n/a

4.1.4 Has the choice of input variables been substantiated and coordinated with the parties involved?

n/a

4.1.5 What measures have been taken to prevent unfair or unjustified bias being created or exacerbated in an AI system?

The core architectural design is a measure against bias. By requiring the model to base its answers exclusively on the full text of user-selected documents, it prevents the LLM from

introducing external knowledge or biases. This grounding of the response in a user-approved context is the primary mitigation strategy.

4.1.6 Can the AI system be used by the intended end users (in other words irrespective of their characteristics, such as age, gender or capacity)?

The web interface has been audited using Google Lighthouse. It has received a 91% web accessibility score (chat interface), aimed at e.g. visual accessibility. The unmet criteria are minor and do not present an obstacle to using the application. To use the AI system, a basic level of office computing is required. It is assumed that any user at KiM (proof-of-concept phase) or the ministry (pilot phase) meets this basic level of capabilities.

4.1.7 Are there stop mechanisms, supervision mechanisms or monitoring mechanisms in place to prevent groups in society from being disproportionately affected by the negative implications of the AI system? Specifically for ILT: a distinction needs to be made here between ondertoezichtstaanden (supervised parties (OTS)) and the rest of society.

n/a

4.2 Accuracy

4.2.1 How will the continuous accuracy of the system be measured and safeguarded?

Currently: Accuracy is primarily safeguarded by the core architectural choice: grounding the LLM's answers in the full text of user-selected documents. This drastically reduces the likelihood of out-of-context or fabricated information. The inclusion of citations allows users to manually verify accuracy against the source material. A process for continuous, automated accuracy measurement is not described. Future: We plan to implement an objective benchmark to function as ground truth (working name: KiM-bench). This allows us to evaluate and monitor the completeness, accuracy and neutrality of answers.

4.2.2 What acceptance criteria have been set up to measure the quality of the input (data) and output (data) of the model?

n/a

4.2.3 Are the acceptance criteria appropriate for the data and the purpose of the AI system?

n/a

4.1.4 How will the output (data) be regularly checked at random and continually monitored for correctness?

Users are encouraged to manually validate received answers in the source texts. They are also encouraged to submit feedback in case of unexpected or incorrect behaviour. This allows us to analyse and mitigate issues.

4.1.5 How will deviations in the output (data) relative to the acceptance criteria be analysed and corrected in a timely fashion?

n/a

4.1.6 What would the results be if alternative models were used?

n/a

4.3 Reliability

4.3.1 Is the AI system reliable?

Currently: The system's process is reliable (search, select, answer). The output from the LLM, however, may have some inherent variability. A low temperature setting is used to minimize the variability of answers.

Future: We plan to implement an objective benchmark to function as ground truth (working name: KiM-bench). This allows us to evaluate and monitor the completeness, accuracy and neutrality of answers.

4.3.2 What are the most important factors that influence the performance of the AI system?

n/a

4.3.3 Is a part of the (sub)dataset excluded from the model's learning process and only used to determine reliability or is the model's reliability calculated by means of cross-validation?

n/a

4.3.4 How has the (hyper)parameter tuning been substantiated and assessed?

n/a

4.4 Technical implementation

4.4.1 How has the AI system been implemented technically?

It is a Python web application using the NiceGUI framework for the UI. The backend logic is contained in api_functions.py and communicates with the OpenAI API for vector search (using OpenAI's Vector Store) and for language model inference. The application is designed to be self-hosted and includes authentication mechanisms.

4.4.2 Has there been consideration of how the AI system fits into the existing technical and system infrastructure and have appropriate measures been taken for its roll-out (if applicable)?

n/a

4.4.3 Describe the system architecture (how do the software components interrelate)?

The system uses a two-stage RAG architecture.

- Search stage: An LLM rewrites the user's query for a vector search against a database of publication chunks. The LLM then summarizes these findings for the user.
- Interaction stage: The user selects documents. The full text of these documents is then used as context for an LLM to generate answers and handle follow-up questions.

The structure.md file provides a complete overview of how the different Python files and UI components interrelate.

4.4.4 Have any specific hardware and software requirements been documented?

n/a

4.4.5 If the application is hosted externally, under what conditions is this happening?

The application is hosted by ODC Noord on government-contracted infrastructure. The AI model is hosted by OpenAI and accessed via API.

4.4.6 How is access to the AI system and its components configured (think of the generic

IT management measures)?

The hosted application can be managed by ODC Noord and KiM colleagues with access to ODC Noord's management API.

4.4.7 How can the AI system interact with other hardware or software (if applicable)?

n/a

4.4.8 How is the logging and monitoring configured?

User conversation history is stored on the server until a conversation is ended (erased/restarted), with user consent. Anonymous usage statistics are collected on a data analytics platform (Umami), also hosted by ODC Noord. These analytics are stored persistently, but do not contain identifiable characteristic of the user or the content of their conversation history.

4.5 Reproducibility

4.5.1 Is the AI system reproducible ? Has a process been set up to measure this?

The LLM's generated output may not be perfectly reproducible in future runs due to the probabilistic nature of the model.

4.5.2 Can output (data) obtained be reconstructed now or in the future (i.e. have previous versions of the model, datasets and conditions been saved by means of version management)?

All changes to the source code, including system prompts, are committed to version control (Gitlab). Input data can theoretically be reconstructed by rolling back the dataset (vector store) to include only publications up until a given data. The LLM used can be rolled back as long as specific versions remain available at OpenAI.

4.5.3 Is it possible to reconstruct the model based on the given parameters and a fixed seed?

n/a

4.5.4 Can the broad outlines of the AI system be reproduced using the documentation?

Yes. The entire application code, including prompts, are open-source.

4.5.5 How will the changes be documented during the system's lifetime?

All changes are committed to version control.

4.6 Explainability

4.6.1 Is the AI system sufficiently explainable and interpretable for the developers?

The underlying LLM is a black box. However, the system's process is highly explainable. The two-stage architecture makes the data flow explicit: a query leads to a list of identified source documents, which are then explicitly used as the sole context for the final answer. This process-level explainability is a core design principle.

4.6.2 In developing the AI system, how has account been taken of the model's explainability, for example for the users?

A welcome/splash screen is shown on first access to the application, explaining the system's functionalities and limitations. Users are also invited to consult the source code or to get in touch if any questions arise.

4.6.3 What technologies have been used to ensure that the AI system is explainable and

why was this technology chosen?

n/a

5. Data governance

5.1 Data quality and integrity

5.1.1 Which training data will be used as input for the algorithm and from which sources do the data originate?

The system does not train a new AI model. It uses a pre-trained model from OpenAI. The data used for the retrieval process is the corpus of public research publications from the KiM Netherlands Institute for Transport Policy Analysis.

5.1.2 How will the data quality be safeguarded?

Data quality is contingent on the quality of the original KiM publications, which is generally considered high. The system enforces a level of quality control by requiring specific metadata attributes for each document (e.g., document_title, authors, date) to ensure proper functioning.

5.1.3 Is the data used necessary for the AI system?

Yes. This application requires the data it is intended to access in order to function.

5.1.4 How are you preventing unintended data duplications?

n/a

5.1.5 Is it possible to update the training and test data when the situation requires it? When will you decide to retrain, temporarily stop or further develop the AI system?

n/a

5.1.6 Does the data meet the assumptions underlying the model?

n/a

5.1.7 How has the input (data) used in the AI system been collected and collated?

Input data is sourced from in-house publications published on the KiM website.

5.1.8 How will the data be labelled?

n/a

5.1.9 What factors (think of limitations in the method of collection, storage, etc.) affect the quality of the input (data)? And what can you do about that?

The input documents (PDFs) are converted and parsed as plain text. Information only contained in images is excluded. This is rare; most information is explicit in the published text. In future, we can assess the possibilities using AI-based parsing libraries to include image data too.

5.1.10 Has the input (data) been assessed for changes that occur during training, testing and evaluation? Also during the use of the algorithm over the course of time?

The input data changes only as new publications are made available. The effect of new publications is explicit, as the application requires the user to select which publications are to be used.

5.1.11 If the output (data) is used as new input, how will the output (data) be stored and checked for correctness and completeness?

n/a

5.1.12 How will you ensure that the output (data) is available in a timely fashion?

n/a

5.2 Privacy and confidentiality

5.2.1 What approaches are being adopted for personal data or confidential data?

The system is designed to operate on publicly available documents and does not use confidential data. It does not process personal data, with the exception of the user's query and conversation history, which is stored at the browser level and only until a conversation is ended (erased/restarted), and can be opted out of. The basic authentication system is separate from user conversation data and does not use personal usernames. Usage statistics are collected anonymously and does not include the user's conversation history.

5.2.2 Does the AI system work with personal data (is the GDPR applicable)? If so, please also complete the following questions. If not, proceed to 'Regarding confidential data'.

No, the system's primary data source consists of publicly available research publications and does not contain personal data. The only potential personal data processed is the user's query and conversation history, for which there is an opt-out, and which is not linked to the authentication system. Usage statistics are collected by the Umami platform, which is GDPR-friendly, and do not include the user's conversation history.

5.2.3 Is the processing of personal data proportional and in line with subsidiarity (use the assessment in Chapter 3 as a basis for this)?

n/a

5.2.4 Can the output of the AI system be tracked back directly or indirectly to individuals (is the GDPR applicable)?

n/a

5.2.5 Have officials been involved, such as the Chief Privacy Officer, Information Security Officer, Chief Information Officer, Privacy Officer, etc.?

n/a

5.2.6 How often is the quality of and necessity for processing personal data evaluated?

n/a

5.2.7 Will confidential data be used or stored?

n/a

5.2.8 How will the security of this information be safeguarded?

n/a

6. Risk management

6.1 Risk prevention

6.1.1 How has the system been tested for appropriate and targeted risk management measures?

We have identified little to no risks that need to be managed, as we are working with public data with a clear, limited scope. General risk management steps include secure hosting on

government infrastructure (ODC-Noord).

A risk exists that using non-European (or non-government) cloud services, such as the OpenAI API, becomes restricted for government applications. If so, the system can be switched to use another LLM provider API, including government-hosted vlam. This step has not yet been taken due to current better capabilities of the OpenAI models regarding context length and reasoning.

6.2 Alternative procedure

6.2.1 What will the plan be in the event of problems with the operation of the AI system?

The alternative procedure is the existing manual process: users would revert to searching for publications on the KiM website using the traditional keyword-based search functionality.

6.2.2 What would be the impact of the system failing?

The system is not expected to be vital to any process. If urgent knowledge input is needed and the system is not available, KiM can always be contacted.

6.2.3 See the calculator example above. What equivalent effect could occur if the AI system is put into service and is this desirable?

n/a

6.3 Information security risks

6.3.1 How are information security risks identified, reduced to an acceptable level and tested (from a technical perspective)?

Access to the application is managed through configurable authentication mechanisms (IP address, username/password, or user-provided API key). As the core AI functionality relies on the OpenAI API, the system inherits its security posture.

6.3.2 How are unauthorised third parties prevented from taking advantage of vulnerabilities in the AI system?

n/a

6.3.3 What would the impact be of third parties having unauthorised access to the source code, data or results of the AI system?

The impact would be minimal, as the source code and data is already public, and any results can be reproduced.

6.3.4 Is it possible for people to take advantage of the fact that an AI system is being used instead of a human decision?

n/a

6.3.5 What is the system for recording who is using the AI system and for how long?

Anonymous usage statistics are collected. The length of sessions (amount of messages) can be tracked, but not linked back to individual users. Limiting access is handled by a basic (password/IP-based) authentication system.

6.3.6 In addition to the standard I&W security measures, have additional measures been taken to secure the AI system?

n/a

7. Accountability

7.1 Transparency towards users

7.1.1 In what way do you provide your end users with an insight into the operation and limitations of the AI system? And are these given sufficient attention for as long as they persist?

The two-stage design provides users with direct insight.

- Operation: Users first see the search results and must actively select the documents that will form the knowledge base for their conversation.
- Limitations: This process makes the scope and limitations of the system explicit – it will only use the selected documents to answer questions.

7.1.2 What role do people play in decisions based on the AI system's input ('humans in the loop') and how do you enable them to play this role?

The user is a human-in-the-loop by design. The system does not automatically generate an answer from retrieved documents. Instead, it presents the search results and requires the user to review and select which documents are relevant. This curation step is mandatory before the interaction stage can begin.

7.1.3 How can the system be monitored and understood by everyone (human oversight)?

Human oversight is facilitated by the system's transparency. By citing its sources and providing direct links to the original publications, users can always verify the AI-generated answer against the source material.

7.2 Communication to parties involved

7.2.1 To what extent are you transparent vis-à-vis different groups of parties involved about the AI systems and in what way?

The system is transparent through its user interface, which makes it clear that a user is interacting with an AI. The documentation states the UI includes disclaimers about AI-generated content. Furthermore, the code is fully open-source, providing maximum transparency into its operation.

7.2.2 Are mechanisms being set up to enable end users to make comments about the system (data, technology, target group, etc.)? And how and when are these validated (analysed and followed up on)?

There's a feedback button in the UI and we have a public issue tracker:
https://gitlab.com/kennisinstituut-voor-mobiliteitsbeleid/kim-explorer/-/issues

7.3.3 Pursuant to the AI Act, does the system need to be included in the algorithm register and/or (for high-risk applications) in the EU database?

- Algorithm register: As a Dutch public authority, KiM will register the system in the algorithm register.
- EU database: No. As a minimal-risk application, it does not need to be registered in the EU database for high-risk AI systems.

7.2.4 Are the end user of and parties involved in the AI system informed that the results are generated by an AI system and what this entails for them?

Yes, there is an explanation on the welcome screen as well as a persistent reminder below the chat screen.

7.2.5 Has a manual been compiled?

A technical manual is available in the form of code documentation in our GitLab repository. Brief

end-user instructions are available on the welcome screen.

7.2.6 What are the potential (psychological) side-effects, such as the risk of confusion, preference or cognitive fatigue in the end user of using the AI system?

n/a

7.2.7 In what way are different groups of parties involved (citizens, colleagues, managers, etc.) given an insight into the different aspects of the AI system? This includes such areas as data use, model or results.

n/a

7.2.8 How have you taken measures to achieve explainability specifically towards the end user?

Explainability for the end user is a central feature of the system's "novel" design. The two-stage workflow is the primary measure. By first presenting the user with a list of relevant source publications for them to review and select, the system makes the exact scope and source of its knowledge base transparent before generating an answer.

7.2.9 Is the system sufficiently transparent to enable deployers to interpret the system's output (data) and use it appropriately?

n/a

7.2.10 Have steps been taken to provide end users with training if necessary?

n/a

7.2.11 How are you ensuring that comments made by parties involved and end users are properly handled internally?

There is a persistent feedback button, and the AI responses encourage users to use it. Feedback submissions are received in a shared mailbox monitored by two colleagues, using a labelling system to ensure each message is handled.

7.2.12 If a party involved wishes to lodge an objection,21 or submit a complaint about the AI system,22 is it clear what steps they should take? The same applies to lodging an appeal.

n/a

7.3 Verifiability

7.3.1 How will the AI system be verified and by whom?

The code and functioning has been tested and verified by multiple KiM colleagues, including by authors reviewing AI-generated answers about their own work.

7.3.2 In what way is accountability provided about the AI system?

Accountability is provided through: • Transparency: The two-stage process shows users exactly what information is being used. • Traceability: Answers include citations and links back to the original source documents. • Open source: The application code is publicly available for inspection.

7.3.3 Who provides the independent audit of the AI system? And in what way?

n/a

7.4 Archiving

7.4.1 How is the input (data) stored?

User input is stored temporarily on the server, linked to a user's browser, until the corresponding user clears their chat session. The user can disable this feature (in which case their message history does not persist across page reloads) so that input data is never stored.

7.4.2 What is the retention period for the input (data)?

n/a

7.4.3 How is the model stored?

The model is hosted commercially by OpenAI.

7.4.4 How is version management arranged?

The application code is available on a public GitLab repository.

7.4.5 What is the retention period for the output (data)?

Output is stored temporarily on the server, linked to a user's browser, until the corresponding user clears their chat session. The user can disable this feature (in which case their message history does not persist across page reloads) so that output data is never stored.