‘Your Ethics Friend’: An AI Assistant for University Ethics Review

Álvaro Menéndez[1], Giulio Ginami[2], Nicu Gurmuzachi[3], and Tim Kramer[3]

Human and ethical aspects of AI
MSc Artificial Intelligence & Engineering systems University of Eindhoven

October 24, 2025

Public Summary

In recent years, universities have placed increased emphasis on ethical compliance in student-led research. At Eindhoven University of Technology (TU/e), all Bachelor and Master students conducting research involving human participants or personal data are required to submit their projects to the Ethical Review Board (ERB). While this is essential to protect participants and ensure research integrity, many students encounter challenges in navigating the process, particularly in estimating the risk level of their research. These challenges often result in delays, resubmissions, and administrative inefficiencies.

The main problem we address is the lack of support for students in assessing the ethical risk level of their research. Many students struggle to interpret the risk assessment criteria used by the ERB, resulting in incomplete or incorrect submissions. This leads to additional review cycles, delays, and extra workload for both students and reviewers. Our objective is to develop a tool that automates the preliminary risk assessment using AI. By analysing the completed ERB form, the tool will predict whether the research falls under low, medium, or high risk, guiding students towards a more accurate and complete application.

Comprehensive Summary

Background and Relevance

The whole process of conducting research involving human subjects in universities is based on a commitment to ethical integrity at its core. This moral adherence is in practice through an obligatory ethical review procedure, which is managed by the Institutional Review Boards (IRBs) or Ethical Review Boards (ERBs) among others. Acquiring ethical approval is not just a matter of fulfilling an official requirement; it is the most crucial point where the welfare, rights, and dignity of people participating in the research are protected. The international ethical principles including the ones laid out in the Belmont Report are the main ones that govern this whole procedure. The Belmont Report mentions the three principles of Respect for Persons, Beneficence, and Justice which are considered the core ethical principles (National Commission for the Protection of Human Subjects of Biomedical and Behavioral research, 1979).

According to Bergstraesser et al. [2], many researchers are insufficiently prepared when it comes to creating ethics applications. Basing on the lack of the experiential knowledge, many students consider the procedures to be complex, time-consuming, and bureaucratic according to McLinton et al. [7]. The point of view is widely shared and accepted in academic literature. Research exposes that a considerable number of student researchers feel unqualified to draw up an ethics proposal, which might then cause them to refrain from asking and thus pursuing the valuable research questions [3]. The focus can easily be turned away from the substantive ethical issues in their research design to getting through a seemingly complicated administrative system.

Applications are often incomplete or inadequately presented as a result of the unpreparedness and the obligation to meet the complicated ERB application requirements, which are a Data Management Plan (DMP), informed consent documents, and a clear description of methodology. Proposals are often rejected or resubmitted, and the primary reasons for that are normally superficial errors that are avoidable. As Bueno et al. (2009) found, returns of the project were mainly a result of insufficient information, a vague methodology, and badly phrased informed consent documents. Similarly, McLinton et al. [7] found that reviewers’ comments contained frequent themes, which meant that the applicants were experiencing similar root issues.

The effects of this inefficiency have a double impact. Firstly, it places a significant burden on the students themselves. Bachelor’s and Master’s projects have tight academic deadlines. It may already take the official processing time six weeks for high-risk research (TU/e) to be approved; if a proposal needs to go through several rounds of reversion, these Backlogs can seriously hamper a student’s chances of finishing their research right on time. Secondly, the number of incomplete or deficient applications places an undue administrative workload on the Ethical Review Boards. As pointed out by Page and Nyeboer [9], time-to-decision has been one of the most reliably vilified components of the ethics review system. When board members are required to cycle back repeatedly to examine the same proposals to identify elementary errors, it siphons their limited time and attention away from being devoted to really difficult ethical issues in high-risk research.

When grasping this context, the tool *“Your ethics friend”*, takes the clear role of a useful assistant. Thereby

not replacing human oversight or control, simply facilitating the ethics review process for students and reviewers, since improved ethics forms will result in easier and thus faster processing times. The quality of an ERB form for example, can be greatly improved by checking for spelling mistakes and possible inconsistencies before diving into the adherence to rubrics given by the ethics committee. In essence, this tool has the potential to alleviate student effort by bringing better clarification and trust into the ethics review process.

Problem Definition and Objective

Students frequently encounter challenges such as uncertainty about what documents are required; misunderstanding of the timelines involved and a lack of early insight into the ethical risk level of their project. These challenges can delay the review process and lead to incomplete submissions, unnecessary communication loops with the committee, and frustration on the part of both students and reviewers. Students lack a clear understanding of how their research project will be evaluated in terms of ethical risk. Consequently, they often fail to submit all the required documentation, or they underestimate the time needed for approval (at TU/e, ethical review process for low/medium risk takes approximately 2 weeks [4], for high risk approximately 6 weeks [4]). This creates delays and inefficiencies for both students and ethics committees.

From a reviewer's point of view, high volumes of applications are submitted around the same timeline. A large percentage of these forms come from students completing the review process the first time, resulting in a high number of revisions for preventable reasons. These circumstances accumulate to a state were a bottleneck is present, extending the timelines for feedback, ultimately pulling more resources from the ERB Page and Nyeboer [9]. This problem distracts ethical review boards from their main purpose; reviewing high risk research, in the end this inefficiency slows down overall academic advancements.

Objectives:

The main goals of *"Your ethics friend"*:

- To develop a tool that can **estimate the ethical/privacy risk level** of a research project based on key input parameters (e.g., data type, participant type, data collection method).
- To provide **personalized guidance** on the documents and steps required.
- To promote **ethical awareness and compliance** among students early in the research design phase.

Plan of Approach / Design Strategy

The key to the success of this project lies in its roots, specifically in adopting a design for values (DFV) approach. Such a methodology means incorporating societal and ethical values in the architecture of our tool, instead of keeping these values as an afterthought [5].

The development of our tool will comprise of three steps:

Step 1: Identifying the relevant values

We do not want our tool to replace any stage of the review process; it is rather intended as a supplementary tool for the student. Our main goal is therefore to focus on the students' trust. To do so, we adopt a Design for Values (DFV) approach, a methodology that means incorporating societal and ethical values in the architecture of our tool, instead of keeping these values as an afterthought. We have explicitly defined these as our main operational values: transparency and explainability, fairness and freedom from bias, and finally privacy.

Step 2: Embedding of Values

Over the past few years it has been demonstrated that LLMs are prone to *hallucinate*, this is, when a generative AI model produces incorrect, nonsensical, or misleading information while presenting it as fact [6].

This is a high risk for our tool, as it would make the tool unreliable and untrustworthy for the users.

A great way to mitigate this risk is to use Retrieval-Augmented-Generation (**RAG**). RAG combines retrieval systems with a generative model to create more accurate and context-aware responses [1]. These systems have shown to highly improve the trustworthiness of Gen AI systems [11].

In order to ensure fairness and freedom from bias, the RAG will be built upon protocols from diverse disciplines (e.g., engineering, social sciences) to ensure the AI is not biased toward a single research methodology.

Finally, we want to guarantee privacy by guaranteeing protection of student research data. This will be achieved through several key architectural and policy decisions:

- The entire system will be deployed on secure university servers, ensuring that sensitive research proposals and data never leave the university's protected network.
- We will enforce a 'zero retention' policy; the system will be designed to process documents in-memory, and all uploaded data will be immediately and permanently destroyed after the analysis is complete.

Step 3: Assessing Success and Evaluation

Ensuring the tool stays aligned with its core values and correct is of utmost importance. Various types of key performance indicators (KPIs) will be used to evaluate success:

Quantitative metrics: In order to test the success of our tool, there are some specific metrics that can be measured:

- **Accuracy on evaluating study risks**: Defining accuracy as the percentage of correct risk estimations (Low, Medium, High) made by the tool when compared to the final classification given by the human ERB reviewers.
- **Reduction in Revision Cycles**: Measure the average number of resubmissions or revisions required for students who used the tool versus those who didnt.

- **Decrease in Time-to-Approval**: Track the average time from a student's initial ERB submission to their final approval. A decrease would indicate the tool is reducing delays.
- **User Adoption Rate**: Calculate the percentage of students who use the tool before their official submission.

Qualitative metrics:

- **Users Opinions:** As detailed in the "Stakeholder Engagement" section, conduct surveys with students to assess the perceived clarity, usefulness, and trustworthiness of the AI's feedback. Do they trust the RAG based responses?
- **ERB Reviewer Feedback (Expert Validation):** Solicit qualitative feedback from ERB members on whether the quality and completeness of applications submitted via the tool have improved, and if it has reduced their administrative workload.

Additionally, a continuous feedback loop will be available to accept and implement feedback from users.

Stakeholder Engagement and Co-Creation

To ensure the tool is responsible and effective, we will adopt a co-creation approach involving all key stakeholders. For end-users, we will conduct surveys with Bachelor and Master students to assess the clarity and perceived utility of the suggestions; their satisfaction will serve as a key qualitative metric. For domain experts, we will engage ERB members to review AI-generated suggestions, ensuring alignment with their review procedures and risk classifications, as their feedback is crucial for validation. Finally, we will collaborate with administrative staff to ensure the tool integrates smoothly into the existing submission workflow, truly reducing the administrative burden.

Social Impact and Responsibility

A primary ethical risk involves the interaction between user trust and tool accuracy. There is a risk of faulty or misleading feedback; although RAG is fundamental to ensure accurate feedback, a misinterpretation of a protocol's could lead to an incorrect risk classification, giving a student a false sense of security. This danger is significantly amplified by the concurrent risk of over-reliance or "de-skilling," as we recognize the risk that students might blindly rely on the tool rather than using it as a guide. To mitigates this combined danger, the tool's design will prioritize transparency and include explicit warnings. Instead of being a 'black box', the tool will be transparent by design, leveraging RAG to ensure every suggestion is verifiable and explains why it produced an output by citing the specific source from the official guidelines it used. Furthermore, to directly counter over-reliance and reinforce that it is "not replacing human oversight or control", the tool will show clear warnings explicitly stating that it is a supplementary assistant, not a substitute for critical thinking, and that the final decision is up to the student and the ERB

Your Ethics Friend benefits students, reviewers, and the university. Students get help fixing the lack of support by being guided to a more accurate and complete application. This saves them effort and builds trust. The Ethical Review Board (ERB) benefits from a reduced extra workload and undue administrative workload. Because they receive improved ethics forms , reviewers get easier and thus faster processing times. This allows them to stop wasting time on elementary errors and focus on really difficult ethical issues in high-risk research. For the university, this creates a more efficient research process, reducing delays and strengthening its commitment to research integrity.

Comparison to Alternatives

An alternative option which would achieve a similar output, as in enhancing transparency and improve the ethics review process for students, would likely be in a workshop format. A traditional, ethics-focused classroom setting where the basics of risk scoring would be explained. This non-technological method would be mandatory for all bachelors and master students conducting research with human subjects, and would provide instruction on completing the ERB application, creating consent forms and understanding the university's ethical guidelines. Furthermore, a tutor would be assigned to this small group of students, to help them with the process and ensure they are following the right path. This tutor would be assigned in addition to the existing supervisor and would review the ethics application before submission. Strong points of this method include human-to-human interactions encouraging rich discussions and debates along with real time feedback. This alternative has key downsides that will hinder its development:

Scalability and Resource Intensity:
For a university with thousands of researchers to create small classrooms with dedicated workshops along with mentor presence and feedback would be resource-intensive. This would pose a real scalability issue for large universities.

Lack of on-demand support:
If a problem occurs on weekends or late at night when working on the ethics review form, immediate guidance could be absent. This crucial 24/7 on demand support is available with *"Your ethics friend"*.

Consistent tutor feedback:
Since the appointed tutors would not necessarily be ERB-approved professors with serious experience. Various answers could be given for sensitive research proposals, however, an AI tool trained on the same principles that the ERB operates by would eliminate any faulty advice.

Precise Feedback:
Lastly, the line-to-line review offered by *"Your ethics friend"* would be far too consuming for tutors or professors in such workshops. These workshops would cover basic principles but would omit precise sentence-based feedback on the student's specific research proposal.

While the hands-on approach does promise a similar outcome, the AI tool was chosen because of concerns such as scalability, consistency and constant accessibility. *"Your ethics friend"* serves as a powerful tool for the ethical re-

view process, by providing personalized, consistent and immediate feedback to students.

Responsible Innovation Project Statement

In order to ensure that technology advancement is in line with society values, ethical norms, and institutional trust, *"Your ethics friend"* was developed in accordance with the principles of Responsible Research and Innovation (RRI). According to Stilgoe, Owen, and Macnaghten [10], the project exemplifies the four RRI dimensions of anticipation, inclusiveness, reflexivity, and responsiveness.

Anticipation

The group gave careful thought to how automating ethical risk assessments can affect society and ethics. Overreliance on algorithmic judgments, possible bias in training data, and misunderstandings of ethical risk levels are among the anticipated problems. A human-in-the-loop technique reduces these concerns by guaranteeing that competent ERB reviewers, not the AI system, retain the final ethical decision.

Inclusiveness

Academic advisers, ethics board reviewers, and students' viewpoints were all taken into consideration during the design process. Feedback loops were incorporated into the project at every stage, allowing stakeholder values like accountability, openness, fairness, and user empowerment to be explicitly reflected in the system requirements.

Reflexivity

The group gave careful thought to the ethical presumptions that underlie the automation of ethics evaluation. The tool serves as a helpful guide rather than a substitute for moral deliberation since this reflective approach recognizes that ethical reasoning cannot be fully formalized. To guarantee that the initiative adhered to the three main tenets of the Belmont Report[8]: respect for persons, beneficence, and justice; internal conversations and ethical check-ins were conducted.

Responsiveness

The design of the system includes feedback integration and ongoing learning processes, enabling adaptability in response to changing institutional policies and ethical standards. Iterative improvements will be informed by user feedback and recurring audits conducted by ERB professionals, guaranteeing that the system stays impartial, transparent, and reliable.

By basing its design on RRI and Design for Values (Friedman and Nissenbaum [5]), the initiative seeks to uphold the integrity of academic research, advance procedural justice, and raise ethical awareness. "Your Ethics Friend" promotes responsible behavior among student researchers and democratizes access to ethical literacy, but it does not take the place of ethical reasoning.

User Narrative

Clara is a Master's student in Industrial Design at TU/e, finalizing the research plan for her thesis. Her project involves analyzing form data where participants have provided feedback on a new product concept.

She is anxious about her project timeline, as she needs to start analyzing the data soon to graduate on time. In her draft Ethical Review Board (ERB) application, she states that she will not anonymize the data, keeping personal identifiers linked to the responses.

Before submitting her documents, Clara opens the '*Your Ethics Friend*' tool on the university's intranet and uploads her draft ERB form. After analyzing the document, '*Your Ethics Friend*' provides interactive feedback:

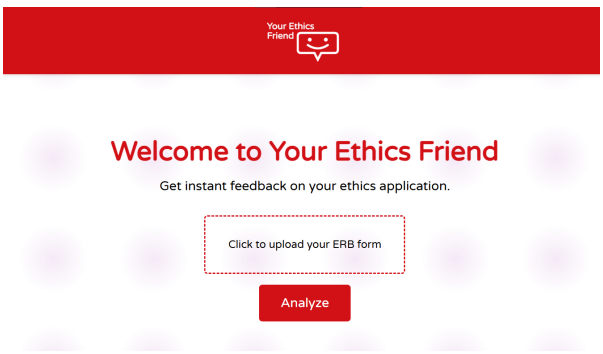
*"Current Plan: In your ERB form you have indicated you will not anonymize the data, therefore with a **99.99%** your study will be classified as **High Risk** by the ERB Comittee .*

Suggested Change: If you fully anonymize the data upon collection, your study can be re-classified as low risk"

Clara immediately sees the value. She hadn't realized that her decision about data handling was the single factor escalating her project to a high-risk category. The tool made the consequence of her choice clear: a four months delay. She realizes that anonymizing the data is a simple change that doesn't compromise her research goals. She immediately revises her application to include a clear anonymization protocol, confident that she will now get approval within two weeks.

Thanks to *"Your ethics friend"*, Clara not only avoided a critical planning error but also learned a key principle of ethical data management. She reduced her uncertainty and can now proceed with her project in a more organized and stress-free manner.

Visual Representation



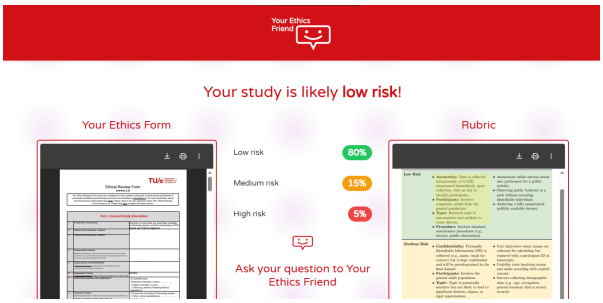


Figure 1: "*Your Ethics Friend*" user interface

References

[1] Amazon Web Services (AWS). *What is Retrieval-Augmented Generation (RAG)?* https://aws.amazon.com/what-is/retrieval-augmented-generation. Accessed: 24 October 2025.

[2] E. Bergstraesser et al. "Deficiencies in paediatric research applications delaying ethics committee approval". In: *Swiss medical weekly* 150 (2020), w20267.

[3] K. Davis et al. "Seeking approval: International higher education students' experiences of applying for human research ethics clearance in Australia". In: *Journal of academic ethics* 20.3 (2022), pp. 421–436.

[4] *Ethical review for Bachelor and Master students*. https://tuenl.sharepoint.com/sites/intranet-ethical-review/SitePages/Ethical-review-for-Bachelor-and-Master-students.aspx. Accessed: 2025-10-24.

[5] B. Friedman and H. Nissenbaum. "Bias in computer systems". In: *ACM Transactions on information systems (TOIS)* 14.3 (1996), pp. 330–347.

[6] Lei Huang et al. "A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions". In: *ACM Transactions on Information Systems* 43.2 (Jan. 2025), pp. 1–55. ISSN: 1558-2868. DOI: 10.1145/3703155. URL: http://dx.doi.org/10.1145/3703155.

[7] S. S. McLinton et al. "Evidence-Based Guidelines for Low-Risk Ethics Applicants: A Qualitative Analysis of the Most Frequent Feedback Made by Human Research Ethics Proposal Reviewers". In: *Journal of Academic Ethics* 22.4 (2024), pp. 735–758.

[8] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Tech. rep. Report No. (OS) 78-0012. Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1978.

[9] S. A. Page and J. Nyeboer. "Improving the process of research ethics review". In: *Research integrity and peer review* 2.1 (2017), p. 14.

[10] Jack Stilgoe, Richard Owen, and Phil Macnaghten. "Developing a framework for responsible innovation". In: *Research Policy* 42.9 (2013), pp. 1568–1580. ISSN: 0048-7333. DOI: 10.1016/j.respol.2013.05.008. URL: https://doi.org/10.1016/j.respol.2013.05.008.

[11] Yujia Zhou et al. *Trustworthiness in Retrieval-Augmented Generation Systems: A Survey*. 2024. arXiv: 2409.10102 [cs.IR]. URL: https://arxiv.org/abs/2409.10102.