SoftwareX 25 (2024) 101601

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original Software Publication

IRB-draft-generator: A generative AI tool to streamline the creation of institutional review board applications

Ryan C. Godwin [a,b,*], Ayesha S. Bryant [a], Brant M. Wagener [a], Timothy J. Ness [a], Jennifer J. DeBerry [a], LaShun L. Horn [a], Shanna H. Graves [a], Ashley C. Archer [c], Ryan L. Melvin [a]

[a] Department of Anesthesiology and Perioperative Medicine, The University of Alabama at Birmingham School of Medicine, United States
[b] Department of Radiology, The University of Alabama at Birmingham School of Medicine, United States
[c] The University of Alabama at Birmingham School of Medicine, United States

ARTICLE INFO

Keywords:
Institutional review board
Generative AI
Large language models

ABSTRACT

The Institutional Review Board (IRB) is fundamental to conducting ethical research involving human subjects. IRB applications require detailed descriptions of the research and specific indications of how the research will be implemented. This can be difficult for inexperienced researchers. Preparing the application is a significant time commitment, even for experienced researchers. In order to lighten the administrative burden on busy clinical professionals, this software application will automatically generate a draft human subject research protocol (the most laborious element of an IRB application) based on responses to a short form. This technology uses generative AI and a custom literature search plug-in to draft the protocol from succinct, user-provided details. User inputs include a brief description of the research, including the hypothesis, inclusion/exclusion criteria, and the study design type (e.g., randomized clinical trial). This tool can expedite the IRB application creation process, provide additional consistency for reviewers, and may reduce clinician researcher burnout through a reduction in clerical work thereby facilitating participation in meaningful research.

Metadata

Nr	Code metadata description	*Please fill in this column*
C1	Current code version	v0.1
C2	Permanent link to code/repository used for this code version	https://github.com/UABPeriopAI/IRB_Assistant
C3	Permanent link to reproducible capsule	https://github.com/UABPeriopAI/IRB_Assistant/releases/tag/v0.1
C4	Legal code license	GPL
C5	Code versioning system used	Git
C6	Software code languages, tools and services used	Python, Streamlit, Docker, OpenAI (e.g. GPT4)
C7	Compilation requirements, operating environments and dependencies	Docker, VSCode, WSL2/Linux
C8	If available, link to developer documentation/manual	https://github.com/UABPeriopAI/IRB_Assistant/tree/gh-pages
C9	Support email for questions	ryangodwin@uabmc.edu

1. Motivation and significance

Institutional Review Boards (IRB) provide essential evaluations of research plans involving human subjects to ensure that research is carried out ethically and in the best interests of the participants. This review of research involving human subjects is critical, as historically, transgressions have been carried out against underrepresented (e.g., racial minorities) or vulnerable (e.g., prisoners) populations in the name of research. Thus, the importance of the IRB application and review process cannot be overstated. Due to the complexity of thoroughly weighing participant ethical considerations, an IRB application must be carefully prepared and reviewed.

Generating quality applications for IRB approval is time-consuming and labor-intensive. The challenging aspects of creating an IRB application can impede early-career investigators and be a time sink for experienced researchers, particularly in relation to the human subject research protocol itself. Many articles have elucidated the IRB process and have highlighted ways to optimize the process [1–3], yet there has been little development of support infrastructure related to the actual

* Corresponding author.
E-mail address: ryangodwin@uabmc.edu (R.C. Godwin).

https://doi.org/10.1016/j.softx.2023.101601
Received 5 September 2023; Received in revised form 16 November 2023; Accepted 27 November 2023
Available online 8 December 2023

writing of the application. Inevitably, IRB applications constructed without proper guidance often wind up in an iterative cycle costing the applicant and the reviewers additional time and effort.

Time is a precious resource for clinicians, particularly for those attempting to conduct research while also managing a full clinical workload. It is notable that academic clinicians have some of the highest rates of burnout for all professionals. Burnout metrics are commonly based on the Maslach Burnout Inventory, which incorporates measures of emotional exhaustion, depersonalization, and low sense of personal accomplishment [4]. Perhaps most importantly, burnout is implicated with worse patient outcomes [5,6].

Burnout is believed to be driven by the heavy demand on the time of clinicians [6,7]. Clerical and regulatory burdens, many of which are related to the Electronic Health Record (EHR), have significantly contributed to the increased rates of physician burnout. Reductions in administrative burden for clinicians reduce burnout rates [8] and while there is no apparent literature directly linking IRB application burden to burnout rates, any reduction in administrative and regulatory efforts for clinicians can potentially have a positive impact. Evidence also suggests that clinicians who perform work they find meaningful have reduced rates of burnout [9]. For some clinicians, research is empowering as it provides an opportunity to help patients at a broader scale but it takes time. For this reason, we believe that a reduction in the time needed to draft an IRB application will both facilitate research and reduce academic clinician burnout.

The recent success of Generative AI, as popularized by ChatGPT (OpenAI Incorporated, San Francisco, CA), is opening new opportunities for natural language AI tools to support healthcare workflows. From summarizing radiology reports [1] to helping a patient better understand a technical diagnostic report [10], there are many ways in which generative AI can have a positive impact on healthcare. Research protocol drafting is a way to utilize generative AI for healthcare applications which minimizes risks involving protected health information. This IRB drafting tool has the potential to increase the volume of valuable clinical research by lowering barriers to clinical research and democratizing the IRB application process for early career clinical researchers.

2. Software description

The IRB draft generation tool expedites the time-consuming process of creating a first draft of a research protocol (the core and most laborious element of the IRB application process). The user provides a short description of the research to be conducted, including the hypothesis, inclusion and exclusion criteria, and the study design type (selected from a drop-down menu) in an easy to use, web-based interface. A workflow schematic of the software is available in Fig. 1 which highlights the four components of user required input, the generative AI core component that accepts the user input, creates the literature search, and incorporates those results into the drafted, downloadable document of the drafted research protocol.

2.1. Software architecture

The IRB draft generator can be deployed with different large language models (LLMs) at the core as there are different trade-offs to consider—particularly between speed and quality. Typically, older models (e.g., GPT-3.5) run faster but with worse quality than newer models like GPT-4. While the software does not currently extend this option to the end user, it is important to note that the tool has an adaptable LLM endpoint that can incorporate improved LLMs in the future. The code uses the open-source Python package LangChain to interface with the LLM. Since LangChain abstracts away the particular LLM used, only one line of code would need to change to support a new or alternative LLM once an interface is available in LangChain.

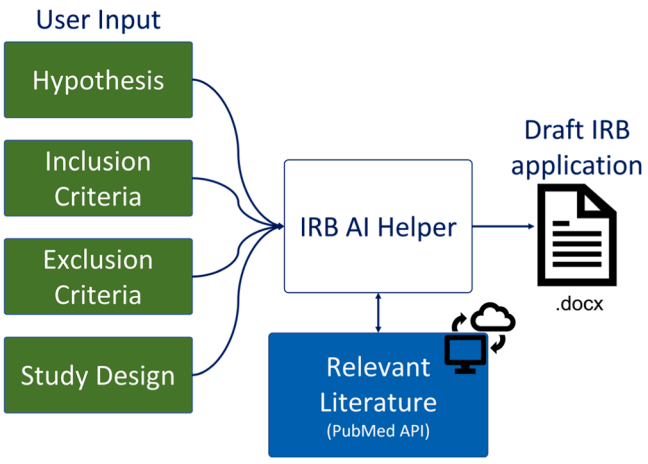

Fig. 1. The Workflow of the IRB Draft Generator includes 4 user inputs, the literature search tool, the generative AI of the IRB Helper and the output document. The IRB Helper uses the user input to generate a relevant literature search query which is fed to a literature search tool that interfaces directly to PubMed.

Software development began from an adapted template for Machine Learning Operations described in previous work.[1] The template provides a springboard for development by providing an operational framework supporting documentation, debugging, package management, and maintenance.

2.2. Software functionalities

2.2.1. Generative AI for drafting

The core functionality of the software is the generative AI drafting component. This essential functionality integrates the user input with the literature search results and a pre-defined template structure. Built to configure different LLMs, the generative AI is intentionally flexible and designed to adapt to the institution's needs.

The drafting tool integrates the user input with configured prompting scripts embedded in the application. This aspect ensures that the LLM has plenty of context awareness that is specific to drafting IRB protocols, but not specific to any research project, allowing the user to worry only about their specific research protocol. The backend context depends on the particular study design requested by the researcher. Just as the user interface updates automatically depending on the study design selected, the backend context updates accordingly as well. This integration provides the LLM the specific context for the user's research topic and expected IRB protocols, but still requires recent literature to be sufficiently comprehensive.

2.2.2. Literature search

The IRB-Drafting application incorporates an essential literature searching capability by integrating directly with an internally built literature search plugin. The literature search is essential as LLMs sometimes hallucinate scientific references [11–13]. A hallucination is when the generative AI creates real-looking citations that do not exist. However, LLMs are adept at using contextual information from text to create reasonable search queries. By integrating into the PubMed application programming interface (API), the application can

[1] Godwin RC, Melvin RL. Toward Efficient Data Science: A Comprehensive MLOps Template for Collaborative Code Development and Automation. *SSRN Electron J*. doi:https://dx.doi.org/10.2139/ssrn.4541529

expediently query and retrieve relevant literature in incorporate context from those articles into the IRB drafting tool. This feature provides a substantive rationale to the protocol draft, based on actual published work (i.e., not hallucinated), for the proposed research. While summaries of the literature search tool may still hallucinate, the references are pulled directly from PubMed and are, therefore, accurate.

2.2.3. Document drafting

The authors include clinical researchers and reviewers of IRB applications who preferred draft output as a Microsoft Word document. As such, the software requires that the AI output conforms by providing a specified markdown format. This markdown gets populated by the configured LLM, and the combined output (the generative AI output plus some boilerplate language) gets converted to a Microsoft Word document for download and review. To download the file, the user pushes the corresponding button (see code documentation for further details: *https://github.com/UABPeriopAI/IRB_Assistant/tree/gh-pages*.

2.2.3.1. Human engineering. The tool incorporates a number of elements that ensure researchers clearly review and revise the IRB draft. First, upon accessing the tool, investigators must acknowledge certain disclaimers, as highlighted in Fig. 3. Then, upon draft completion, the first page of the output is a checklist of actions for the researcher to carry out prior to submission with general tasks that include:

- Add PI and Co-I names
- Add coordinator, assistant, and or resident names
- Confirm all investigators are current with IRB and GCP training
- Confirm the list of variables to be collected is exhaustive for your study
- Add sample-size information as appropriate
- Add study budget to the resources section if applicable
- Note if you want assistance adding your study to clinicaltrials.gov
- Review and edit all sections

An additional, study specific checklist is provided depending on the type of study as there are different considerations for a retrospective study versus a prospective interventional study. With these checks in place an IRB will be able to easily identify any applications that were not revised by the investigator(s).

2.2.4. Text simplification

Reviewers on an IRB who evaluate the proposed research typically come from different educational backgrounds. For example, in the United States, an IRB must have at least five members and represent diverse perspectives. Therefore, IRBs mandate that while there needs to be representation from a scientific background, there must also be representation from individuals whose focus is primarily non-scientific. To accommodate the backgrounds of all reviewers, the LLM generates text that can accommodate a non-technical audience. This feature ensures that individuals can understand the research protocol without specific domain knowledge in the research area, and reviewers can focus on the ethical considerations rather than understanding the proposal's verbiage.

3. Illustrative examples

A screenshot sample shows the primary web-based user interface in Fig. 2. It highlights the instructions for filling out the form along with user inputs, including the research question, results of the literature search, and the inclusion criteria. Additionally, it shows an example of input for a research project where the investigators want to explore differences in pain management for patients who underwent regional vs. general anesthesia.

The user inputs all of the required information and simply hit the "*Draft Document*" button. The application provides additional contexts in the back-end to properly conform the output to IRB standards. Once the LLM completes the text generation the user interface provides a button that allows the user to download the document.

The final output provides a draft of the research protocol that contains human-engineered features to ensure the user acknowledges and understands the risks of using generative AI. To that end, the first page of

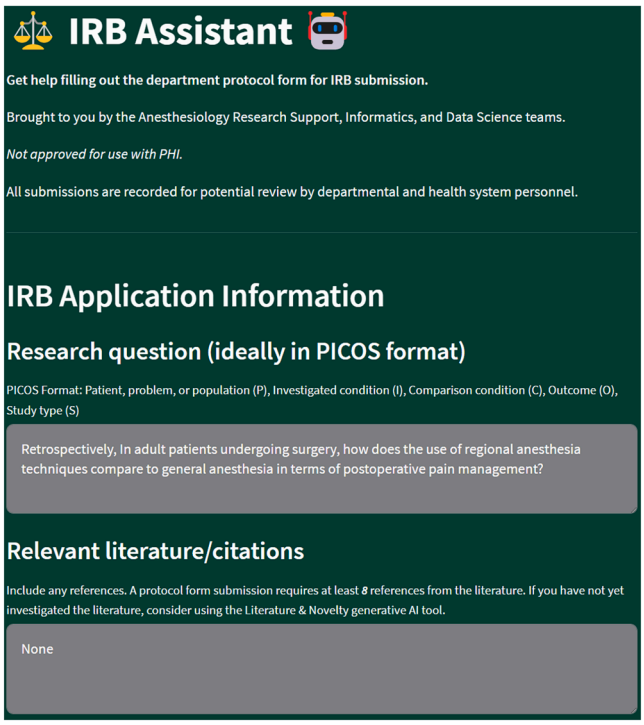

Fig. 2. The user interface for the IRB draft generator. The web-based interface provides instructions to fill out the form and dialog boxes for the user to input their information. After submission the generative AI creates the draft and provides a downloadable document to the user.

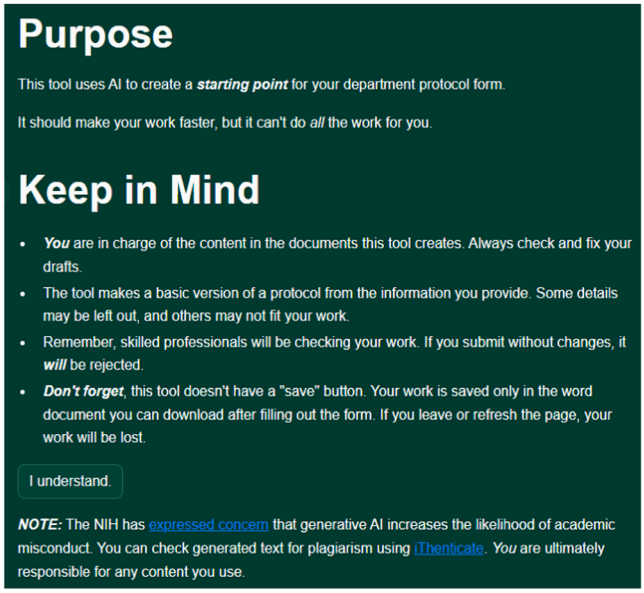

Fig. 3. Application Disclaimer. Before accessing the tool, investigators must acknowledge that it is for drafting purposes only and that they are ultimately responsible for the submitted IRB application.

the output draft provides a checklist for the user to complete before submission. This human-in-the-loop aspect of the tool is intentional, requiring researchers to review and approve all text generated by the LLM. The drafted IRB will require revision, and this step ensures that the researcher has performed those revisions and that they take responsibility for the contents of the application.

A sample of the checklist items includes:

1. Add PI and Co-I names;
2. Confirm that the list of variables to be collected is exhaustive for your study. Changing this later may require IRB amendments or reimbursements (i.e., repeating the process);
3. Add sample-size information (including any study groupings) as appropriate;
4. Review and edit all sections.
5. These checklist items are placed at the top of the output document to ensure that the researcher reads this before getting to the content of the IRB application. Additionally, this procedural requirement helps application reviewers identify proposals submitted without review and revision by the investigator.

4. Impact

The potential impact of this tool is profound. The most obvious impact is on the immediate time saving offered to the researcher in generating an initial draft, as this is often the most time-intensive portion of the process. Furthermore, there is emerging evidence that human-AI collaborations can outperform either alone in healthcare settings [14,15]. In addition to time saving and quality improvements, this tool standardizes the IRB process so that, with broad adoption, protocols for IRB applications will have the same formatting and structure at scale, bringing a consistency to the applications that can help IRB application reviewers be more efficient in their evaluation.

This work can also bolster the quality of the research plan by allowing researchers more time to dedicate to planning and preparing their research. A recent article in *Science* suggested that generative AI improved the turnaround times of workers by 40 % overall and that the quality of the work increased by 18 % [16]. To realize these gains for drafting IRB applications would provide extra time for clinicians to connect with their patients, develop their research (e.g., grants, manuscripts, experiments), or take wellness respite. This, in turn could reduce burnout as it reduces professional time commitments by creating a reasonable first draft for an IRB application.

Evaluating the efficiency and quality improvements offered by the IRB drafting tool is planned future research. In the meantime, to ensure human subject protocol quality, a departmental research team reviews all protocols before submission to the IRB (both human and human with AI drafts). The authors recommend others implementing this tool incorporate similar additional review of IRB applications until the tool is validated.

There are potential downsides to this application of generative AI if it is not used judiciously. Without due diligence in review and revision from both researchers and the IRB, complications could emerge, as there is no guarantee that the draft is factually correct throughout. Accordingly, we've integrated key principles of the Swiss cheese model of patient safety into our approach. This model, developed by James Reason, employs multiple layers of safety measures - each representing a barrier to potential errors [17,18]. In our context, these layers consist of checklists, acknowledgements of limitations, and varying levels of scrutiny depending on the nature of the research. While this doesn't provide a guarantee against all errors, it greatly reduces the risk by making any oversight clear and certain misuses (e.g., not revising the draft) readily rejectable (see Section 2.3.1).

Another important consideration when using generative AI in sensitive applications like clinical research, is to know where the information is sent. In our deployment, the IRB drafting tool is protected by a Business Associate Agreement with the vendor that covers use of sensitive data. Users interested in deploying the tool for use at their institution will need to implement their own LLM endpoint (the open-source code has a non-functional, mock endpoint) and in so doing, take care to protect their sensitive data and ensure institutional evaluation and approval of the endpoint before deployment. By implementing these layered checks, we aim to reduce potential harm and align the tool with responsible AI usage in clinical research.

5. Conclusions

The IRB-Draft-Generator represents a significant step forward in leveraging generative AI in health research administration. By simplifying and expediting the creation of IRB applications, the tool saves time and resources and introduces additional rigor and comprehensiveness to the process. Despite its early success, generative AI is not a panacea and has limitations. The IRB protocol drafting tool incorporates human engineering elements to ensure that the tool is always with critical human review. It also fills a critical need to help researchers adhere to essential IRB regulations more quickly and effectively than was previously possible.

Declaration of generative AI

During the preparation of this work the authors used ChatGPT (GPT4) to outline and draft sub-sections of the manuscript and document code. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of Competing Interest

Ryan Godwin co-owns multiple LLCs (Lotz, Lacasse, Godwin, LLC - a control system software company and Blaze Pascal – a consulting firm) neither of which have any interests or connections to this work.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Code available in public repository

Acknowledgments

None.

References

[1] Shoenbill K, Song Y, Cobb NL, Drezner MK, Mendonca EA. IRB process improvements: a machine learning analysis. J Clin Transl Sci 2017;1(3):176–83. https://doi.org/10.1017/cts.2016.25.
[2] Lynch HF. Opening closed doors: promoting IRB transparency. J Law, Med Ethics 2018;46(1):145–58. https://doi.org/10.1177/1073110518766028.
[3] Pech C, Cob N, Cejka JT. Understanding institutional review boards: practical guidance to the IRB review process. Nutr Clin Pract 2007;22(6):618–28. https://doi.org/10.1177/0115426507022006618.
[4] Maslach C, Jackson SE. The measurement of experienced burnout. J Organ Behav 1981;2(2):99–113. https://doi.org/10.1002/job.4030020205.
[5] Edú-valsania S, Laguía A, Moriano JA. Burnout: a review of theory and measurement. Int J Environ Res Public Health 2022;19(3). https://doi.org/10.3390/ijerph19031780.
[6] Rotenstein LS, Brown R, Sinsky C, Linzer M. The association of work overload with burnout and intent to leave the job across the healthcare workforce during COVID-19. J Gen Intern Med 2023;38(8):1920–7. https://doi.org/10.1007/s11606-023-08153-z.
[7] Harry E, et al. Physician task load and the risk of burnout among US physicians in a national survey. Joint Comm J Qual Patient Saf 2021;47(2):76–85. https://doi.org/10.1016/j.jcjq.2020.09.011.

[8] Shanafelt TD, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. Mayo Clin Proc 2016;91(7):836–48. https://doi.org/10.1016/j.mayocp.2016.05.007.
[9] Shanafelt TD, et al. Career fit and burnout among academic faculty. Arch Intern Med 2009;169(10):990–5. https://doi.org/10.1001/archinternmed.2009.70.
[10] Moor M, et al. Foundation models for generalist medical artificial intelligence. Nature 2023;616(7956):259–65. https://doi.org/10.1038/s41586-023-05881-4.
[11] Masters K. Medical Teacher's first ChatGPT's referencing hallucinations: lessons for editors, reviewers, and teachers. Med Teach 2023;45(7):673–5. https://doi.org/10.1080/0142159X.2023.2208731.
[12] Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. Nat Med 2023:2–4. https://doi.org/10.1038/s41591-023-02412-6.
[13] Mcgowan A, et al. ChatGPT and Bard exhibit spontaneous citation fabrication during psychiatry literature search. Psychiatry Res 2023:115334. https://doi.org/10.1016/j.psychres.2023.115334.
[14] Reverberi C, et al. Experimental evidence of effective human-AI collaboration in medical decision-making. Sci Rep 2022;12(1):1–10. https://doi.org/10.1038/s41598-022-18751-2.
[15] Cabitza F, Campagner A, Sconfienza LM. Studying human-AI collaboration protocols: the case of the Kasparov's law in radiological double reading. Heal Inf Sci Syst 2021;9(1):1–20. https://doi.org/10.1007/s13755-021-00138-8.
[16] Noy S, Zhang W. Experimental evidence on the productivity effects of generative artificial intelligence. Science 2023;381(6654):187–92. https://doi.org/10.1126/science.adh2586.
[17] Reason J. Human error: models and management. West J Med 2000;172(6):393–6. https://doi.org/10.1136/ewjm.172.6.393.
[18] Wiegmann DA, Wood LJ, Cohen TN, Shappell SA. Understanding the 'Swiss Cheese Model' and its application to patient safety. J Patient Saf 2022;18(2):119–23. https://doi.org/10.1097/PTS.0000000000000810.