Review Article

The Moral Responsibility of Designers Regarding AI: Protocol for a Meta-Aggregative Systematic Review

Cássio Henrique Bauer[1], Magnus Ferreira de Melo[2], Flávio Anthero Nunes Vianna dos Santos[1]

1. Universidade do Estado de Santa Catarina, Brazil; 2. Universidade Federal de Santa Catarina, Brazil

This study presents a systematic review protocol to explore the moral responsibility of designers in the context of artificial intelligence (AI). The review synthesizes qualitative evidence to address the ethical implications of designer choices and their impact on the development, improvement, and use of AI systems, products, and services. By leveraging a meta-aggregative approach, the study examines key questions: What constitutes the moral responsibility of designers? How do ethical concerns in AI implicate designers, and what frameworks or considerations can guide their responsibilities? The protocol outlines a rigorous methodological process, including database searches, grey literature analysis, citation tracking, and quality assessments, employing tools such as EndNote, Rayyan, and the JBI Critical Appraisal Checklist. The outcomes aim to inform actionable ethical frameworks, enhance design practices, and contribute to the broader discourse on responsible AI development. This research seeks to provide normative recommendations and clarify how designers can engage with ethical considerations to foster a morally robust vision for AI-integrated systems.

1. Background

Designers have a moral and social responsibility. Papanek[1] had already argued that the interventions made by designers shape what is and still will be, but they only happen when they accept their moral and social responsibilities. This works because design is a social practice capable of producing social transformations, and not the designer who holds this power or is solely responsible for the adverse effects. Thus, to say that designers have a moral responsibility is possible when one understands this activity as socially situated. Furthermore, as the characteristics of designers can be evidenced

 doi.org/10.32388/9USOEO

throughout the design process and its outcome, it is valid to ensure that the intended use can potentially differ from the initial intention, leading to a social or moral responsibility.

The focus of this work, on the other hand, is to think about moral responsibility through the moral values incorporated into designs. Buchanan[2] explained, for example, that designers create persuasive arguments that emerge when users think about or use a product. To construct them, designers draw on technological reasoning, the way in which materials and processes are manipulated; character, designs reflect their creators; and emotion, what is emotionally desirable and valuable. It is therefore credible to say that designers inscribe their morality in the things they create, since artefacts have affordances for action[3] and are basically the designer's choices. Or even that designers materialise morality[4].

The moral responsibility of designers tends to be approached as an aspect that arises from the interaction between individuals and technologies. Verbeek[4] suggests that it can be identified in technological mediation, a theory that states that human-world interactions are mediated by technologies (human-technology-world). For him, the designer's moral responsibility includes the unintended use of designed technologies, when such uses could have been minimised or resolved beforehand.

When we think about it in conjunction with recent developments in artificial intelligence, things change. The task of identifying it is difficult, as the behaviour of these systems can vary considerably from the intentions of their designers. Not only that, but artificial intelligence systems also tend to be exempt from moral responsibility due to their design[5], falling on the designer or user[6].

Considering the lack of studies that have applied a systematic review (with meta-aggregation) about designer moral responsibility and analysing its relationship with artificial intelligence systems, this research aims to synthesise qualitative evidence on the topic of the designer's moral responsibility, considering its relationship with artificial intelligence. Specifically, its secondary objectives are (a) to identify what the designer's moral responsibilities are, (b) to identify ethical issues about artificial intelligence that implicate the designer, (c) to relate the moral responsibility of designers to ethical issues about artificial intelligence, (d) to explain the designer's moral responsibility to artificial intelligence, and (e) to demonstrate how this is implicated in the development/improvement and use of artificial intelligence systems/products/services.

It is hoped that the results of the meta-aggregation will serve as direct recommendations for practice involving the moral responsibility of designers, by investigating opportunities for building a positive and normatively strong moral vision[7], as well as bringing more clarity to the development/improvement of ethical models of action.

2. Research questions

2.1. Primary research question

What is the designer's moral responsibility in relation to AI? How is it configured regarding the development/improvement and use of AI systems/products/services?

2.2. Secondary research questions

What is the moral responsibility of designers? What questions about AI raise the moral responsibility of designers? What are the ways of considering the moral responsibility of designers in the development, improvement, and use of AI systems, products, and/or services?

3. Review team

As this work is part of a master's thesis, the student responsible and his supervisor are integrated into the team. In addition, another individual has joined the team to help with the process of sorting the relevant documents. As a result, the team (advisory group) is listed below with their respective CRediT assignments:

- Cássio Henrique Bauer: Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.
- Magnus Ferreira de Melo: Formal analysis, Investigation, Data Curation, Visualization.
- Flávio Anthero Nunes Vianna dos Santos: Validation, Supervision, Data Curation.

4. Type of review

This research is of a basic nature, dealing with qualitative data. When considering its objective, it is classified as an exploratory study, since it aims to identify the elements that reinforce the designer's moral responsibility. As such, it employs procedures relating to systematic review strategies with

meta-aggregation, which follows the proposal of Pettricrew and Roberts (2006) for the Social Sciences, plus the recommendations of Lockwood, Munn, and Porritt[8] to meta-aggregation. To draw up the protocol, the suggestions of Moher et al.[9] were applied. However, unlike what they recommend, the current protocol does not meet the PROSPERO requirements and should contain some health-related outcome involved in the research, which is not the case. For this reason, registration and publication will be carried out in the Open Science Framework (OSF), using the protocol template file developed by Cruz-Martínez[10].

5. Review stages

The methodological outline of this work is as follows: (1) check for the existence of other systematic reviews on the topic of interest; (2) draw up the guiding research question; (3) define the team (advisory group); (4) draw up the research protocol; (5) search phase; (6) screening phase; (7) methodological quality assessment phase; (8) data extraction phase; (9) analysis and synthesis phase; (10) evidence confidence analysis phase; and (11) report writing phase.

6. Search strategy

6.1. Databases

To find relevant documents, a search will be carried out in databases, namely: Academic Search Premier - ASP (EBSCO), PsycINFO (APA), SAGE Journals Online, ScienceDirect (Elsevier), Scopus, WOS, SocINDEX, arXiv, BASE: Bielefeld Academic Search Engine, ERIC, IEEE Xplore, OpenAlex, Lens.org, philpapers, and ACM DigitalLibrary. Many of the databases were chosen based on the study by Chapman[11]. Databases powered by artificial intelligence were included due to the abundance of bibliographic material they can retrieve, which is advantageous for productions around the area of Design and nearby fields, given their plurality of indexing in other areas and in the databases themselves (for example, a journal specialising in Design can be indexed in several categories in Scopus), which proves to be a challenge.

6.2. Grey literature

Grey literature will be included in the study, given the intended scope of this type of study. There will be no distinction between the other types of literature and this one, as the studies included in this

review will be analysed equally, even though the evaluation of the evidence will be applied. That said, the main source of information used for collection will be Google Scholar.

6.3. Other search strategies or resources

Other methods of searching for relevant documents include citation and journal searches. The citation search will use eligible documents and will be carried out via citationchaser. Thus, the following procedures will have to be carried out[12]: firstly, identifying the group of documents (the eligible documents); secondly, importing the unique identifiers for each document (DOI); thirdly, downloading the references and citations in a RIS file; fourthly, starting the activities of screening these documents. The search for journals, in turn, will be manual and carried out as follows: with all the relevant documents included after the citation search, the journals in which they were published will be listed and, from there, a search will be carried out using the strings used previously, or dividing them into smaller parts. A document containing all the information about the manual search carried out, given its flexibility, should be attached to the work to allow reproducibility by other researchers.

6.4. Inclusion and exclusion criteria

The eligibility of the studies to be analysed was defined according to the Population, Phenomenon of Interest, and Context (PICo) framework, namely: designers (population), moral responsibility (phenomenon of interest), and interaction with artificial intelligence (context). Thus, the study will include theoretical, conceptual, empirical articles, book chapters, conference papers, and/or essays of a qualitative nature; published in English or Portuguese; in which designers are the population under study; in which moral responsibility (or harm, concern, implication, action, consequence, liability) is addressed or in which the ethical issue (or benefit, concern, project, consideration, or principle) of AI is addressed. From these, we will exclude those that: are research of the type of literature review (scoping, mapping, literature review, for example), (full) books, metric studies (bibliometrics, scientometrics, among others), quantitative method research, letters to the editor, editorials, retraction, inaccessibility of the full text; or which do not relate the designers to the phenomenon of interest or the context.

6.5. Search strings or queries

The search strategies are divided into three stages: pre-collection (applied before the final search for refinement), active collection (use of the refined strategies), and post-collection (search for documents via citations and journals of the material included in the review). Search expiration and repetition will not be considered, as there is no justification for this in the review topics.

The pre-collection strategies are described below. Guiding concepts (1): key terms are defined (designer, moral, responsibility, AI). Correlated concepts (2): using the databases (Scopus, WOS, and Lens.org) with the strings (title, abstract, and keywords: designer ADN moral* AND respons*), the results obtained in each are exported in CSV (Scopus), TXT (WOS), and RIS (Lens.org) formats. Individually, each downloaded file is imported into VOSViewer using the tool that allows you to create a map based on bibliographic data. The type of analysis applied is that of co-occurrence (min. frequency of 5), with 'all keywords' as the unit of analysis, and 'full counting'. Thus, possible correlated terms are identified by subjective analysis by the authors. String refinement (3): correlated concepts are added to the strings, such as harm, consequence, accountability, and implication (terms correlated with responsibility). Based on the results, the need to break down the main research question into secondary questions was noted. Breakdown of strings (4): based on the results of the previous stage, specific strings are created for each of the secondary questions. The last stage of this pre-collection, to be carried out after the protocol has been registered and published, concerns the validation of the strings (5): the strings will be evaluated according to the parameters of quasi-sensitivity and precision (quasi-gold standard), while the databases will be evaluated according to the number of unique relevant documents retrieved in each of them[13].

The method by which validation via quasi-gold standard occurs is described below. Identification of publication sources (1): the criteria for selecting sources are subjective; therefore, Scopus, Web of Science, and Lens.org will be used. Construction of the quasi-gold standard (2): by using the disassembled strings, the search results from each database must be screened (title and abstract) based on inclusion/exclusion criteria. The eligible documents constitute the quasi-gold standard (n =?). Definition of strings (3): Based on word frequency (titles and abstracts), two new strings must be constructed. Search in sources (4): Searches across all databases are conducted using the disassembled strings and the new strings, and the documents must be screened (titles and abstracts). Evaluation of search performance (5): For each database, the quasi-sensitivity parameter will be calculated (number of eligible documents in the database / number of documents included in the quasi-gold standard X

100). Precision is calculated by dividing the number of eligible documents in a database by the total number of documents retrieved from it X 100. If the quasi-sensitivity value is between 72% and 80% and the precision value is between 15% and 25%, the string will be considered valid. Additionally, if the quasi-sensitivity value exceeds 80%, it will also be considered valid, regardless of the precision value. The results of the disassembled and new strings will be compared. If the parameter values are unacceptable, the number of publication sources will be increased, and the process will be repeated until acceptable levels are reached. Databases without unique relevant documents will be disregarded as they would not retrieve potential results[14].

Active collection strategies encompass the validated search activities: search in previously evaluated databases (1) and search in grey literature databases (2), which, in this case, will be Google Scholar; for this purpose, the Publish or Perish tool will be used.

Finally, post-collection activities refer to search activities that should occur after the methodological quality assessment of the studies. These tasks are described in detail in another section.

Database[1]	What is the moral responsibility of designers?	What questions about AI raise the moral responsibility of designers?	Filters
Scopus	designer AND moral* AND (respons* OR account*) designer W/3 moral* W/3 (respons* OR account*)	designer AND (ethic* OR “ethic* design” OR "ethic* princip*" OR "ethic* consider*") AND (AI OR "artific* intellig*”) designer W/3 (ethic* OR “ethic* design” OR "ethic* princip*" OR "ethic* consider*") W/3 (AI OR "artific* intellig*”)	Article titles, abstracts, and keywords Date: 2019 to 2024 Language: English or Portuguese
WOS	designer AND moral* AND (respons* OR account*) designer NEAR/3 moral* NEAR/3 (respons* OR account*)	designer AND (ethic* OR “ethic* design” OR "ethic* princip*" OR "ethic* consider*") AND (AI OR "artific* intellig*”) designer NEAR/3 (ethic* OR “ethic* design” OR "ethic* princip*" OR "ethic* consider*") NEAR/3 (AI OR "artific* intellig*”)	

[1] *Search strategies for other databases will be included in the final report.*

7. Screening

7.1. Screening stages

The screening process will begin with the importation of the results obtained during data collection into EndNote for deduplication. Subsequently, the first round involves the evaluation of titles and abstracts, conducted independently by two of the study's authors. In the second stage, a full-text reading of the documents will be carried out, once again applying the inclusion/exclusion criteria. All bibliographic fields will remain visible during the eligibility check. Any disagreements between reviewers will be resolved through (1) discussion between the reviewers and, if there are difficulties in reaching a consensus, a third reviewer should be called in. The documents identified during post-

collection will undergo screening, following the same principles. Screening instructions have not been drawn up, as the team is small and discussion among themselves is encouraged.

To guarantee the reliability of the screenings, a few measures were taken. The first is that, given that two authors of the review will be screening the documents, reviews will be created in the Rayyan tool, and all the references will be screened by each of them. The reviewers' decisions will be analysed in the Statistical Package for the Social Sciences (SPSS) using Cohen's kappa test, where the acceptable value must be between 0.60 and 0.79 (strong agreement level) or higher (McHugh, 2012).

The procedures are intended to increase the transparency and confidence of the scientific community in the proposed review process. In addition, much of the research in Design is in the Applied Social Sciences, which demands systematisation in certain methodological designs, but this is not what is observed in certain studies. The tools available were included in the study to increase the validity of this systematic review with meta-aggregation.

All the data from each stage, task, and decision should be published in the ‘figshare’ repository, as it is open access. However, this data should only be published after the journal has accepted or authorised the dissemination of the results.

7.2. Screening tools

The data will be handled using the EndNote software (version 20.4.0.16272 for Microsoft Windows) for deduplication of duplicate records. The Rayyan (web-based) tool will be used for eligibility verification as well as for recording decisions.

8. Data extraction

8.1. Data items

The items to be extracted from the qualitative sources included in the review are in line with the proposal by Lockwood et al.[8]. Descriptive information will be extracted: Reference (Last name, First name, Year), Publication date (Date), Context (As reported in the text), Study type (Design), Phenomena of interest (As reported in the text), Exposure (As reported in the text), Source (Journal), Data collection type (Collection strategy), Data analysis type (Collection strategy). Additionally, findings will be extracted: Finding (Text), Illustration from the publication (Direct quotation), Plausibility level (Unequivocal, Dubious, or Unsupported). And condition of extraction: Whether the

data extraction is complete (Yes/No). Any modifications made to the extraction form must be reported in the review report.

8.2. Extraction stages

The data extraction process will consist of two phases. In the first phase (pilot extraction), data equivalent to 25% of the database (n =?) will be extracted regarding descriptive information to refine the extraction form. If new fields are added to the form, they should be incorporated, and the extraction should be redone. If no further changes are necessary, the pilot extraction phase should be concluded. In the second phase, if no further refinement is needed, the complete extraction of descriptive information will be performed. Both phases will be conducted manually by two authors. Extractor instructions have not been drawn up, as the team is small and discussion among themselves is encouraged. This review will not implement procedures or tests to ensure that the extractions achieve acceptable levels of reliability. The reconciliation extraction process will be carried out through (1) discussion between the reviewers and, if there are difficulties in reaching a consensus, a third reviewer should be called in. The extraction of findings will take place during the analysis and synthesis phase. All the data from each stage, task, and decision should be published in the 'figshare' repository, as it is open access. However, this data should only be published after the journal has accepted or authorised the dissemination of the results.

8.3. Extraction tools

The extraction form is shared between the two extractors and will be manually completed in a spreadsheet using Microsoft Excel (version 2411 for Windows).

9. Quality assessment

The methodological quality assessment of the eligible qualitative documents will be verified using the Critical Appraisal Checklist for Qualitative Research, developed by the Joanna Briggs Institute (JBI). As a result, to be included in the review, each article must achieve a 70% "yes" response rate, with each "yes" corresponding to 10% of the total. In any case, two authors will independently conduct the assessments and then must reach a common agreement. Once completed, the post-collection search should be conducted.

10. Synthesis

10.1. Synthesis method

First, a descriptive analysis of the characteristics of the identified studies will be presented, including publication date, data collection type, and study type. Additionally, the synthesis of findings will be conducted through meta-aggregation with the support of the thematic analysis framework by Braun and Clarke[15].

10.2. Synthesis procedure

The first step involves extracting the findings, as well as the illustration and plausibility level. This will be achieved through repeated readings of the documents included in the study[8]. In operational terms, this means reading and rereading the database; systematically coding relevant aspects of the data; gathering extracts for each created code[15], with each extract being accompanied by an illustration; and the plausibility level being defined for each extract. Moreover, each document must have a finding, hence an extract. The findings, illustration, and plausibility level should be included in the extraction form.

Once the findings for each work have been identified, they will be grouped based on the similarity of concepts present among them, resulting in the development of a key concept and an explanatory statement: categories[8]. This means grouping the codes into themes; combining all the data related to a theme; checking if the themes align with the extracts and codes[15]; and each theme (category) is reduced to a key concept and an explanatory statement.

Finally, for each group of categories, a synthesized finding should be elaborated in the form of an indicative statement[8]. In other words, this means creating a thematic map; refining the themes; generating names and definitions for each theme[15]; and elaborating an indicative statement.

10.3. Confidence in evidence

Each synthesized finding will be evaluated using the ConQual tool, which assesses the dependability and credibility of the syntheses (Munn et al., 2014). Qualitative evidence begins the procedure with a high pre-classification (Pearson et al., 2024). Dependability is evaluated through 5 questions with yes/no answers, where 4-5 "yes" answers mean the article remains unchanged; 2-3 "yes" answers

mean the article drops one level; and 0-1 "yes" answers mean it drops two levels. Regarding credibility, the synthesized findings are classified as unequivocal, credible, or unsupported. After this step is completed, the summary table of findings should be filled out. The analysis of confidence in the evidence will be carried out jointly by everyone on the team, with the aim of reaching a consensus.

Appendix A - JBI Critical Appraisal Checklist for Qualitative

Item	Reviewer:			Date:	
	Author:	Year: 1254		Record n.º:	
		Yes	No	Unclear	Not applicable
1	Is there congruity between the stated philosophical perspective and the research methodology?				
2	Is there congruity between the research methodology and the research question or objectives?				
3	Is there congruity between the research methodology and the methods used to collect data?				
4	Is there congruity between the research methodology and the representation and analysis of data?				
5	Is there congruity between the research methodology and the interpretation of results?				
6	Is there a statement locating the researcher culturally or theoretically?				
7	Is the influence of the researcher on the research, and vice- versa, addressed?				
8	Are participants, and their voices, adequately represented?				
9	Is the research ethical according to current criteria or, for recent studies, is there evidence of ethical approval by an appropriate body?				
10	Do the conclusions drawn in the research report flow from the analysis, or interpretation, of the data?				

doi.org/10.32388/9USOEO

Item	Reviewer:			Date:	
	Author:		Year: 1254	Record n.º:	
		Yes	No	Unclear	Not applicable
Overall appraisal:	Include:	Exclude:	Seek further info:		
Comments (Including reason for exclusion):					

Appendix B - ConQual Summary of Findings Example

Systematic review title: **Population:** **Phenomena of interest:** **Context:**					
Synthesized Finding	**Type of research**	**Dependability**	**Credibility**	**ConQual Score**	**Comments**

Notes

The guarantor of the review: Cássio Henrique Bauer.

Acknowledgements

This study is funded in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Funding Code 001.

References

1. ^PAPANEK, Victor. Arquitectura e Design: ecologia e etica. Lisboa: Edições 70, 1995.
2. ^Buchanan R (1985). "Declaration by Design: rhetoric, argument, and demonstration in design practic e." Design Issues. 2 (1): 4. doi:10.2307/1511524.
3. ^JELSMA, Jaap. Designing 'moralized' products: theory and practice. In: VERBEEK, Peter-Paul; SLOB, A driaan (ed.). User Behavior and Technology Development: shaping sustainable relations between consu mers and technologies. Dordrecht: Springer, 2006. p. 221-232.
4. a, bVERBEEK, Peter-Paul. Moralizing technology: understanding and designing the morality of things. C hicago: The University Of Chicago Press, 2011.
5. ^Courtenage S (2023). "Intelligent machines, collectives, and moral responsibility." AI and Ethics. 4 (2): 485-498. doi:10.1007/s43681-023-00285-6.
6. ^IDSKES, J. A model of autonomy for artificial agents. 2016. 57 f. Dissertação (Mestrado) - Curso de Phil osophy Of Science, Technology And Society, Faculty Of Behavioural, Management, And Social Sciences, University Of Twente, Enschede, 2016.
7. ^Donia J, Shaw JA (2021). "Ethics and Values in Design: a structured review and theoretical critique." Sci ence and Engineering Ethics. 27 (5): 1-32. doi:10.1007/s11948-021-00329-2.
8. a, b, c, d, eLockwood C, Munn Z, Porritt K (2015). "Qualitative research synthesis." International Journal of Evidence-Based Healthcare. 13 (3): 179-187. doi:10.1097/xeb.0000000000000062.
9. ^Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA (2015). "Pref erred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement." Systematic Reviews. 4 (1): 1-9. doi:10.1186/2046-4053-4-1.
10. ^CRUZ-MARTÍNEZ, Roberto Rafael. Template for a Systematic Literature Review Protocol. 2023. Dispo nível em: https://research.utwente.nl/en/datasets/template-for-a-systematic-literature-review-proto col. Acesso em: 03 jan. 2025.
11. ^Chapman K (2021). "Characteristics of systematic reviews in the social sciences." The Journal of Acade mic Librarianship. 47 (5): 102396. doi:10.1016/j.acalib.2021.102396.

12. ^Haddaway NR, Grainger MJ, Gray CT (2022). "Citationchaser: a tool for transparent and efficient forward and backward citation chasing in systematic searching." Research Synthesis Methods. 13 (4): 533-545. doi:10.1002/jrsm.1563.
13. ^Zhang H, Babar MA, Tell P (2011). "Identifying relevant studies in software engineering." Information and Software Technology. 53 (6): 625-637. doi:10.1016/j.infsof.2010.12.010.
14. ^Campbell A, Taylor B, Bates J, O'Connor-Bones U (2017). "Developing and Applying a Protocol for a Systematic Review in the Social Sciences." New Review of Academic Librarianship. 24 (1): 1-22. doi:10.1080/13614533.2017.1281827.
15. a, b, c, dBraun V, Clarke V (2012). "Thematic analysis." APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological. p. 57-71. doi:10.1037/13620-004.

Declarations

Funding: This study is funded in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Funding Code 001.

Potential competing interests: No potential competing interests to declare.