

What Makes Code Generation Ethically Sourced?

Zhuolin Xu

zhuolin.xu@mail.concordia.ca

Concordia University

Montreal, Canada

Qiushi Li

qiushi.li@mail.concordia.ca

Concordia University

Montreal, Canada

Chenglin Li

chenglin.li@mail.concordia.ca

Concordia University

Montreal, Canada

Shin Hwei Tan*

shinhwei.tan@concordia.ca

Concordia University

Montreal, Canada

Abstract

Several code generation models have been proposed to help reduce time and effort in solving software-related tasks. To ensure responsible AI, there are growing interests over various ethical issues (e.g., unclear licensing, privacy, fairness, and environment impact). These studies have the overarching goal of ensuring ethically sourced generation, which has gained growing attention in speech synthesis and image generation. In this paper, we introduce the novel notion of *Ethically Sourced Code Generation* (ES-CODEGEN) to refer to managing all processes involved in code generation model development from data collection to post-deployment via ethical and sustainable practices. To build a taxonomy of ES-CODEGEN, we perform a two-phase literature review where we reviewed 803 papers across various domains and specific to AI-based code generation. We identified 71 relevant papers with 10 initial dimensions of ES-CODEGEN. To refine our dimensions and gain insights on consequences of ES-CODEGEN, we surveyed 32 practitioners, which include six developers who submitted GitHub issues to opt-out from the Stack dataset (these *impacted users* have real-world experience of ethically sourced issues in code generation models). The results lead to 11 dimensions of ES-CODEGEN with a new dimension on code quality as practitioners have noted its importance. We also identified consequences, artifacts, and stages relevant to ES-CODEGEN. Our post-survey reflection showed that most practitioners tended to ignore society-related dimensions despite their importance. Most practitioners either agreed or strongly agreed that our survey help improve their understanding of ES-CODEGEN. Our study calls for attention of various ethical issues towards ES-CODEGEN.

CCS Concepts

- Software and its engineering → Software maintenance tools;
- Human-centered computing → Empirical studies in collaborative and social computing.

Keywords

Ethical Sourcing, Responsible AI, Ethical Code Generation

* Corresponding Author.



This work is licensed under a Creative Commons Attribution 4.0 International License.

ICSE '26, Rio de Janeiro, Brazil

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2025-3/26/04

<https://doi.org/10.1145/3744916.3773192>

1 Introduction

Due to the rising demand in automated code generation to reduce software development and maintenance costs, companies like Meta and OpenAI have proposed several code generation models (e.g., Code Llama and Codex) for solving software engineering tasks (e.g., automated program repair [53, 60, 66] and program synthesis [10]). However, with the advent of responsible AI, there are rising global concerns about various ethical dimensions of AI models to ensure ethical and sustainable development. First, practitioners are hesitant to use code generation models due to unclear licensing information and potential copyright infringements. For example, an Intercept lawsuit [1] claimed that “According to the award-winning website Copyleaks, nearly 60% of the responses provided by Defendants’ GPT-3.5 product in a study conducted by Copyleaks contained some form of plagiarized content, and over 45% contained text that was identical to pre-existing content”, indicating the severity of copyright infringements in generated code. Second, several companies (e.g. Apple [93] and Samsung [28]) have banned the usage of AI-based code generation models due to privacy concerns over data leakage. Third, several ethical guidelines [39, 43] were introduced to govern and guide the design of responsible AI models by emphasizing various ethical principles (e.g., transparency, justice and fairness, non-maleficence, responsibility, and privacy). Fourth, there exist emerging needs to consider the environmental impact of AI by characterizing the carbon footprint of AI computing [84, 92, 96]. With the overarching goal of ensuring *ethically-sourced* generation, these initiatives encompass various ethical dimensions that practitioners need to consider when developing and adopting code generation models for software development and maintenance.

“Ethically sourced” refers to “managing all processes of supplying the firm with required materials and services from a set of suppliers in an ethical and socially responsible manner” [47]. In a broader view, ethical sourcing is usually associated with the concept *triple bottom line* that stresses the importance of measuring success in three key factors: people (social impact), planet (environmental impact), and profit (financial performance) [31]. Meanwhile, the core of ethical production involves *supply chain management*, ensuring that each point of the supply chain is obtained ethically and sustainably. In the context of AI systems, the idea of AIBOM (AI Bills of Material) [97, 98] were introduced to describe all the components of an AI system to ensure AI transparency [94].

The concept of ethical sourcing has been adopted in various domains (e.g., economy [75] and medicine [45]) but is still a new concept for the generative AI community. For speech synthesis [85],



Figure 1: Supply Chain of ES-CODEGEN (Stages and Artifacts)

the ethical questions on voice donation (e.g., voice ownership and identity) were discussed. For image generation, Adobe released Firefly, an image generator, in March 2023 [6], claiming to use solely legal and ethical sources for training, making it less likely to induce copyright issues, and “commercially safe”. However, there remains controversial debates on its ethical claim as a Bloomberg report [21] revealed that 5% of its training data were not being ethically sourced as they were from another image generator. As CodeGen models are at the early stage of adoption, it is important and timely to study the ethically sourced issues for code generation.

In this paper, to the best of our knowledge, we for the first time introduce the notion of *Ethically Sourced Code Generation* (ES-CODEGEN) to refer to managing all processes involved in code generation model development from data collection to post-deployment and artifacts involved via ethical and sustainable practices throughout the supply chain as shown in Figure 1. Prior ethically sourced research [47] revealed four main research areas that were widely discussed (i.e., socially responsible supply management, unethical purchasing behaviour, ethical sourcing codes of conduct, fair and ethical trade). Do software practitioners also care about the same set of dimensions when defining ES-CODEGEN? Considering the unique characteristics of AI-driven code generation pipeline [96], what are the key software artifacts and stages that contribute to ethical code generation supply chain management? Prior studies show that unethical sourcing may lead to negative impacts such as loss of reputation [46] and legal risks [85]. Are these impacts any different in the case of unethically sourced code generation (UnES-CODEGEN)? As the first study to understand the various dimensions of ES-CODEGEN and consequences of UnES-CODEGEN, our study aims to answer the questions below:

- RQ1:** What are the dimensions that define ES-CODEGEN from the practitioners’ perspectives?
- RQ2:** Which stages and what kinds of artifacts constitute an ethical supply chain management in the context of code generation?
- RQ3:** What are the potential consequences of unethically sourced code generation (UnES-CODEGEN)?
- RQ4:** What are the acceptable trade-offs and contamination when enforcing ES-CODEGEN?

Our study aims to raise awareness by proposing the concept and contributing to the body of knowledge about ES-CODEGEN. To understand the scope and various dimensions of “ethically sourced”, we first conduct a systematic literature review of prior studies on it. Our interdisciplinary literature review covers publications across various domains, and lays the foundation towards a taxonomy of dimensions involved in defining ES-CODEGEN.

Based on our literature review, we conduct a survey questionnaire involving 32 practitioners. Specifically, the surveys targeted software practitioners familiar with code generation, contributors of code generation models, and researchers in code generation. Our

study expands our understanding of the various dimensions that define ES-CODEGEN, and potential consequences of UnES-CODEGEN, identifies areas that research and practice on ES-CODEGEN support should focus on, and provides a discussion of potential solutions to overcome these consequences.

In summary, this paper makes the following contributions:

Systematic Literature Review: We perform a two-phase systematic literature review using a combination of database search, forward and backward snowballing to adopt a multidimensional view in defining *ethically sourced code generation* (ES-CODEGEN). We first identified various dimensions of ethical sourcing from various domains, and then performed a domain-specific literature search. We manually reviewed 803 papers and identified 71 relevant papers. Our review identified 10 dimensions that define ES-CODEGEN.

Survey: We surveyed 32 practitioners to shed light on practitioners’ views on ES-CODEGEN, including their definitions of ES-CODEGEN, evaluation on provided dimensions, the applicability of dimensions to stages and artifacts in the supply chain of ES-CODEGEN, the dimensions that they tend to ignore in defining ES-CODEGEN, potential consequences of UnES-CODEGEN, and acceptable trade-offs.

Discussion: We discuss the gaps within existing code generation models and ES-CODEGEN models, and highlight the key areas that researchers and practitioners should focus on to build ES-CODEGEN.

Evaluation: We also evaluated the usefulness of our survey. In the post-survey reflection, most participants either agreed (56.3%) or strongly agreed (18.8%) that their understanding of ES-CODEGEN has improved after taking our survey, indicating that our survey helped raise awareness of the importance of ES-CODEGEN. Our anonymized survey questionnaire and results are publicly available at: <https://github.com/xuzhi2021/ESCodeGen>.

In summary, we noted the key findings below from our study:

- ES-CODEGEN is defined by 11 important dimensions: (1) Subject Rights, (2) Equity, (3) Access, (4) Accountability, (5) Intellectual Property (IP) Rights, (6) Integrity, (7) Code Quality, (8) Social Responsibility, (9) Social Acceptability, (10) Labor Rights, (11) Environmental Sustainability. Among these dimensions, stakeholders were most concerned about (1), (5), and (11), and tended to ignore (8), (9), and (10) despite their importance.
- Developers who spent time opting-out from the training data of the Stack dataset [19] (*impacted users*) noted the importance of opt-in consent as a key dimension of ES-CODEGEN.
- All stages and artifacts in ES-CODEGEN’s supply chain are considered to be relevant to the dimensions.
- Participants have listed many consequences of UnES-CODEGEN, e.g., exploitation of developers, low-quality code, monopolization of generative AI, and impact on open-source communities. Among the given options, most stakeholders cared about lawsuit issues, security, and privacy risks.
- When considering various trade-offs in ensuring ES-CODEGEN, most participants noted that accuracy loss is unacceptable (most can accept up to 10% accuracy loss).
- Most participants thought that none of the existing code generation (CodeGen) models aligns with the definition of ES-CODEGEN or only partially aligns (need to improve on transparency and opt-in consent), indicating a considerable gap between current models and ES-CODEGEN.

2 Background and Related Work

We will discuss most of relevant papers in our literature review in Section 3. In this section, we focus on two lines of research closely related to our work: (1) approaches related to ES-CODEGEN, (2) empirical studies on ethically sourced issues.

Ethically Sourced Code Generation (ES-CODEGEN). Recent trends in open-source code generation models focus on various dimensions towards ethical generation, including enhancing AI transparency [72, 77], green code generation [91, 92], bias evaluation of code generation (CodeGen) [32, 58]. To enhance transparency of training data, the Stack dataset [77] contains source code from permissive licenses, and its data governance plan allows developers who do not wish to have their code used for pretraining large language models to opt-out. The Stack dataset is used to train many code generation models (e.g., StarCoder [55] and OLMo2 [72]).

Studies on ES-CODEGEN. Most studies focus on understanding ethically sourced issues in various domains [23, 47], and some focus on the software supply chain [97, 98]. The most closely related work to our work is the literature review of ethical sourcing [47] (published in July 2018), which combines a systematic literature review and a citation network analysis to understand the relationship between the reviewed literature. Although we rely on a literature review to derive the dimensions defining ES-CODEGEN, our work differs in that: (1) instead of citation network analysis, we perform a two-phase hierarchical review to obtain diverse perspectives of ethically sourced dimensions, (2) our focus on code generation offers a technical view of ethically sourced issues, which is absent in prior work, (3) compared to prior study published on 2018, our study offers an updated view of ethically sourced issues.

3 Systematic Literature Review

To obtain a comprehensive definition of ES-CODEGEN, we conducted a two-phase hierarchical systematic literature review (broad → specific), using a combination of database searching, forward and backward snowballing [42]. Figure 2 shows the overall workflow of our literature review. To obtain a broad and high-level view of ethical sourcing from multiple disciplines, we first conducted a survey of research across various domains (e.g., economy [75], medicine [45], and religion [7]) as there is a lack of a clear definition of ES-CODEGEN. In the second phase, we focused specifically on LLM-based code generation literature and annotated the relevant dimensions to the development of ES-CODEGEN.

3.1 Cross-disciplinary General Literature Search

3.1.1 Initial Database Search. We searched for “ethically sourced” on Google Scholar to obtain a comprehensive definition and related dimensions across various domains. Since this term is not specific to the computer science field, we did not use specialized databases such as IEEE Xplore or ACM Digital Library. The search returned about 112,000 results, which were too many to review in full, so we manually examined the top 100 papers sorted by relevance. After excluding non-scholarly publications (55 papers remaining) and selecting relevant ones, we obtained 45 papers. Among the top 100, we found that after the top 51 papers, no new dimensions emerged. Based on this review, we established the synonyms of

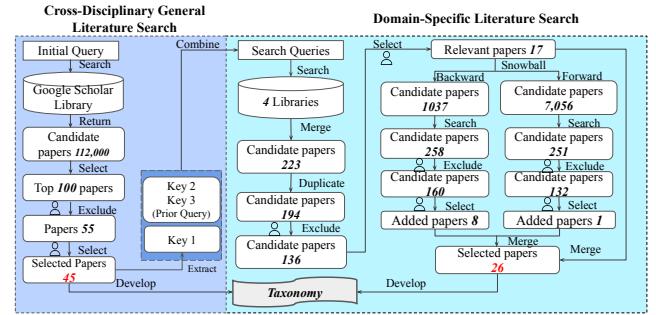


Figure 2: The Workflow of Our Literature Review

Table 1: Database Search Queries and Scopes

Database	Query Keywords Summary	Search Scope
IEEE	key1' AND key2 AND key3	All Metadata
ACM	key1' AND key2 AND key3	Abstract
Scopus	key1' AND key2 AND key3	Article Title, Abstract, Keywords
Google Scholar	key1 * key2 * key3	Anywhere (Default)

“ethically sourced” as the first set of keywords for the domain-specific literature review.

3.1.2 Constructing Keywords for Queries. We prepared three sets of keywords (key_1 , key_2 , key_3) to design the queries for the domain-specific literature review. From the general literature search, we collected key_1 , synonyms of “ethically sourced”. Then, we reused the derived synonyms of “artificial intelligence” and “code generation” from prior work [69] as key_2 and key_3 . Our summarized keyword sets are as follows:

key_1 : ethically sourced, ethical sourcing, ethically acquired, ethically procured, ethical acquisition, responsible sourcing, responsibly sourced, sustainable sourcing

key_1' : ethic* OR responsibl* OR sustainabl*

key_2 : LLM, AI, artificial intelligence, LLMs, large language models

key_3 : code generation, code creation, generate code, code writing, code production, code correction, code quality

3.2 Domain-Specific Literature Search

3.2.1 Query Design and Database Search. Table 1 shows the queries and scopes for four popular databases in the field of computer science: IEEE, ACM, Scopus, and Google Scholar. Initially, we found that queries in the form of key_1 AND key_2 AND key_3 returned zero results in IEEE, ACM, and Scopus, so we simplified and broadened key_1 to key_1' using wildcard operators. For example, we used “ethic*” to cover multiple forms such as “ethics”, “ethical”, and “ethically”. As the length of our query (362 characters without spaces) exceeded the limit for Google Scholar (256 characters), and Google Scholar does not support wildcard operators, we searched each combination of key_1 , key_2 , and key_3 individually, then collected the results, merged and deduplicated them. Since the three databases provide different scope options, we chose a different but appropriate scope for each to include as many results as possible. The query in Google Scholar was finished on March 12, 2025. The queries in IEEE, ACM, and SCOPUS were completed on March 14, 2025.

3.2.2 Merging, Deduplication and Exclusion. We obtained 61 papers from Google Scholar, 79 from IEEE Xplore, 19 from ACM Digital Library, and 64 from Scopus individually, leading to a total of 223 papers. After merging the list of papers and deduplication, we obtained 194 papers. Then, we manually excluded papers that are either not in English or not published in scholarly journals or conferences, and eventually obtained 136 papers.

3.2.3 Paper Selection. As shown in Figure 2, we performed paper selection in several rounds (on 55, 136, 160, and 132 papers). Specifically, we manually excluded irrelevant papers that 1) only reflect on the keywords in sections like “Discussion” or “Related Work” instead of discussing the keyword itself, and 2) superficially mention the keywords, but do not focus on ethical sourcing issues.

3.2.4 Snowballing. We conducted forward and backward snowballing on the 17 relevant papers. This process yielded 7,056 papers through forward snowballing and 1,037 papers through backward snowballing. We then applied keyword filtering to remove irrelevant papers, resulting in 251 and 258 papers, respectively. After manual exclusion, we retained 132 forward and 160 backward papers that were English academic publications. In the last step, we identified one additional relevant paper from forward snowballing and eight papers from backward snowballing.

Combined with the original 17 papers, this resulted in 26 relevant papers. In total, we manually reviewed 100 (general search) + 194 (specific search) + 251 (forward snowballing) + 258 (backward snowballing) = 803 papers, and got 71 (45+26) relevant papers.

3.3 Construction of the Taxonomy

We recorded candidate themes when reading the 71 relevant papers to build a taxonomy with various dimensions that define ES-CODEGEN. Specifically, we followed the recommended steps of thematic analysis [27], and assessed inter-rater disagreement at two key steps. 1) Extract data: Two authors (coders) independently extracted dimension-relevant excerpts for all these papers. The percentage agreement was nearly 90%. A codebook defining various dimensions was also developed in this process and finalized through discussions. 2) Identify and code relevant dimensions: Both coders independently coded all extracted excerpts using the codebook. We transformed the multi-label coding task into a label-level binary matrix, treating each (excerpt, label) pair as a unit, and computed Cohen’s Kappa coefficient [26], and average Jaccard Index [87] per excerpt to reflect annotation set-level consistency. In initial coding (first 10% of excerpts), the Cohen’s Kappa was 0.62 (moderate agreement). After a calibration meeting to refine code definitions, subsequent coding’s Kappa value achieved 0.91 (almost perfect agreement), with the post-calibration average Jaccard Index 0.84 (relatively high similarity). Steps (3)-(5) involved discussions to translate codes into themes, create higher-order themes, and assess trustworthiness. These were performed via multiple meetings among the authors to ensure consistency and validity.

Table 2 shows a structured taxonomy resulting from our thematic analysis, reflecting key dimensions of ES-CODEGEN through three hierarchical levels: (1) **Area**, representing broad areas of ethical concern; (2) **Dimension**, grouping related issues within each

dimension; and (3) **Aspect**, capturing concrete and specific ethical aspects, each with a brief description and relevant citations from the two-phase literature review. This structure allows for both high-level reasoning and fine-grained analysis of ethical sourcing practices in the development of ES-CODEGEN.

3.3.1 Source. This area encompasses the dimensions related to the source of ES-CODEGEN, including subject rights, equity, access, accountability, intellectual property rights, and integrity.

Subject Rights. This refers to whether data subjects’ rights on controlling the use of their source are respected, including:

Informed Consent. In most domains (e.g., biomedical field gathering stem cells [45], and voice generation gathering voice[38]), the standard mechanism for obtaining consent is opt-in, where data subjects are explicitly asked whether they agree to the use of their data for specific purposes. However, CodeGen models often rely on large-scale datasets mined from the web (e.g., GitHub), making obtaining individual consent impractical. To address this, developers of the Stack dataset [19] introduced an opt-out mechanism to allow developers to check if their data has been included and to request its removal from the training. However, developers who chose to opt-out pointed out that the mechanism might still be illegal and morally wrong [18]. To obtain real-world experience of ethically sourced issues in CodeGen, we invited these developers (known as *impacted users* in Section 5.1) to our survey.

Privacy and Security. Prior work [51] mentioned some famous examples like Samsung employees accidentally disclosing sensitive company data to ChatGPT [28], and companies like Google explicitly warning employees not to enter confidential information into chatbots [29]. To address this, some techniques like CodexLeaks [70] are proposed to test privacy leaks of CodeGen models.

Equity. This refers to whether the model development reflects principles of diversity, fairness, and representativeness.

Diversity. Several studies [63, 73, 74] highlighted the importance of diversity in the source of models. Diversity is also emphasized in domains such as medical AI, as prior work [73] noted that ethical AI development is approached through interdisciplinary collaboration and the early involvement of diverse teams across the AI lifecycle.

Fairness. Several studies revealed that the code generated by ChatGPT and Codex exhibits biases related to demographic factors and stereotypes based on gender, race, emotions, class, naming structures, and other characteristics [32, 58]. Several techniques focus on evaluating the bias of CodeGen models [57, 59].

Representativeness. Several studies discussed this aspect to avoid inequity [16, 52, 73, 79, 85]. For example, for voice datasets collection, prior study [15] stressed the need for targeted recruitment and oversampling from underrepresented groups. For CodeGen, another work [52] stated that university programming instructors raised several equity-related concerns, including whether the training data behind AI tools sufficiently represents diverse groups.

Access. This refers to whether resources used in the development of models and models themselves are accessible or properly controlled.

Accessibility. Recent work like StarCoder [56] stressed that closed-access models pose ethical challenges by preventing public oversight, hindering safety assessments, and limiting community contributions. Meanwhile, balancing between privacy and accessibility should also be considered when resources like training data contain

Table 2: Ethics Dimension Taxonomy for ES-CODEGEN (refer to our website for the complete list of references)

Area	Dimension	Aspect	Description	Domain-specific Papers	General Papers
Source	Subject Rights (D1)	Informed Consent (Opt-in [*])	Data subjects are fully informed and give consent to source usages	[56, 63, 74, 76]	[15–17, 38, 45]
		Privacy and Security	Both inputs and outputs avoid private or confidential information	[30, 34, 37, 51, 70]	[15–17]
	Equity (D2)	Diversity	Diverse perspectives (e.g., geographic) are considered	[25, 56, 63, 67, 74]	[33, 35, 38, 44, 73]
		Fairness	Both inputs and outputs are free from biases	[9, 32, 57–59]	[33, 35, 38, 71, 73]
	Access (D3)	Representativeness	Proportional coverage of different groups are considered	[14, 32, 52, 67]	[15, 16, 73, 79, 85]
		Accessibility	Resources are available under appropriate conditions	[37, 56, 76]	[16, 38, 44, 45, 79]
	Accountability (D4)	Transparency	The inputs and development processes are visible	[61, 63, 67, 74, 95]	[40, 46, 64, 73, 82]
		Source Acknowledgement	Sources are properly acknowledged and credited	[51, 54, 61, 63]	None
	IP Rights (D5)	Licensing	The use of inputs complies with relevant licenses	[14, 52, 58, 61, 63]	[44, 85]
		Generation Distinctiveness	The outputs are sufficiently distinct from the origins	[54, 61, 76, 95, 99]	None
Society	Integrity (D6)	Contamination	Inputs are verified and remain free from contamination	None	[7, 11]
		Accuracy+	The generated code is accurate	[25, 58, 63, 74, 95]	[15, 17]
	Social Responsibility (D8)	Community Development	Communities involved in or affected by the development benefit from it	None	[8, 46, 64, 78, 81]
		Social Acceptability (D9)	Religious beliefs, practices, and sensitivities are respected	[25, 56, 59]	[7, 11, 20, 68]
	Labor Rights* (D10)	Culture	Diverse cultural identities and values are considered	[25, 63]	[15, 20]
		Fair Wages*	Human workers involved (e.g., authors of training data*) receive fair pay	[52]	[5, 11, 38, 46, 47]
		Working Conditions	Mentally and physically safe conditions are provided	[52]	[5, 13, 35, 46, 64]
Environment	Sustainability (D11)	Legal Employment	The labor of vulnerable communities is not exploited illegally	None	[5, 35, 46, 64, 75]
		Energy Consumption	The amount of energy and resources used is considered	[14, 25, 86, 91, 92]	[15, 16, 22, 35, 41]
	Emissions and Pollution	The environmental pollution (e.g., carbon emissions) is considered	[25, 52, 86, 92]	[8, 22, 35, 41, 64]	

1 The “Source” area encompasses dimensions focusing on model and data governance

2 The superscript + means that the dimension is a new technical dimension derived from the result of the survey

3 The superscript * denotes that the refined dimension where survey participants noted compensation to authors or input training data is desirable.

4 The superscript ' denotes that the refined aspect where survey participants noted that opt-in informed consent is preferred over opt-out.

sensitive or identifiable information [79]. Moreover, ensuring that AI models are not prohibitively expensive or restricted to certain regions or populations (equitable access) is also important [16].

Accountability. Source accountability refers to whether the development of CodeGen models ensures responsibility and traceability throughout the process, including:

Transparency. The importance of supply chain transparency is widely emphasized across domains [40, 46, 82]. In the context of AI, developers or organizations should disclose what data is used and how outputs are produced to ensure input transparency and output explainability [95]. Several regulatory frameworks (e.g., the GDPR for data processing ¹, and the European Commission’s AI Act for deployment ²) stressed the importance of transparency across different stages of AI systems. Moreover, transparency regarding data sources is important in preventing the misuse of protected or copyrighted content [63].

Intellectual Property (IP) Rights. This concerns whether the development of CodeGen models respects IP rights related to their inputs.

Source Acknowledgement. Although related to transparency, this aspect specifically emphasized the responsibility to provide clear attribution, allowing users to trace the origin of code components to enhance traceability [54, 63].

Licensing. Although many open-source repositories are accessible, many come with specific license terms on permissible uses (e.g., commercial use). However, studies have highlighted that many models have been trained on such repositories without adequately considering or adhering to these licensing terms, raising concerns

about legal violations [14, 52, 58, 61, 63]. For example, the GitHub Copilot investigation raised concerns on potential violation of open-source licenses on AI-generated code [24].

Generation Distinctiveness. Since models can learn and memorize details from the training data, there is a risk that generated outputs may closely mimic or even replicate code in training dataset, raising concerns about compliance and infringement [99]. Hence, developers should ensure generated code maintains adequate distinctiveness from code in training data [54, 95]. Otherwise, the users who deploy generated code that is too similar to its source may unintentionally commit plagiarism [76].

Source Integrity. This considers whether all inputs are sourced from verified sources that comply with ethical standards, with no contamination involved during model development.

Contamination. This aspect is inspired by studies on ingredient evaluation and Halal certification [7, 11], which stated the need of source raw materials strictly from certified suppliers to prevent cross-contamination from non-Halal ingredients.

3.3.2 Society. The social area encompasses social responsibility, social acceptability, and labor rights.

Social Responsibility. This concerns whether the organization or developers of CodeGen models take on broader social responsibilities [8, 46], including:

Community Development. This refers to whether communities involved in or affected by the development of CodeGen models benefit from it, rather than being treated solely as sources of data or labor. For example, Starbucks showed community commitment by adhering to fair trade principles, collaborating with coffee-producing communities, and organizing initiatives (e.g., donations) [46].

Social Acceptability. This refers to if the development process of CodeGen models respects different cultures and religious beliefs.

Philosophy and Religion. This aspect concerns whether philosophical concerns, religious beliefs and sensitivities are respected

¹GDPR: Principles relating to processing of personal data (2023). Available at: <https://gdpr-info.eu/art-5-gdpr/>. Accessed: 2025-06-04

²European Commission (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>. Accessed: 2025-06-04

during model development. For example, prior work [68] discussed ethical concerns raised by certain religious communities regarding COVID-19 antibody vaccines that utilize reagents derived from aborted fetal cell lines. Similarly, for CodeGen, developers should ensure that inputs, outputs, and overall development processes align with commonly held religious principles and avoid content that conflicts with religious sensitivities.

Culture. For example, body donor programs have emphasized integrating cultural and religious needs into operational practices [20]. Prior study [15] also mentioned that trustworthy voice AI should respect cultural sensitivity throughout data collection, synthetic data generation, and implementation. These principles about culture are equally applicable to CodeGen models.

Labor Rights. This dimension concerns whether organizations or developers respect the basic labor and human rights of workers involved in the development processes of CodeGen models.

Fair Wages. For example, a prior study [52] cited a news³ that OpenAI underpaid the Kenyan data labeler (less than \$2/hour), raising concerns about fair pay in the global AI supply chains.

Working Conditions. For example, reducing a model's toxicity often requires human labelers to handle harmful content, which may affect their mental health. In the previous news³, workers reported that the wellness services provided by the organization were limited and ineffective under the strain of heavy workloads.

Legal Employment. For example, previous research [46] cited a report [100] revealing child labor in global brands such as Nike, which demonstrated the importance of this aspect.

3.3.3 Environment. This area examines whether the development of CodeGen models aligns with principles of sustainability.

Sustainability. This dimension assesses whether the development of models is conducted in an environmentally sustainable manner.

Energy Consumption. Several studies [86, 92] have raised concerns about the high energy cost of LLMs and proposed approaches to evaluate energy efficiency (e.g., data center energy efficiency, hardware efficiency, and algorithmic efficiency). However, the generated code of existing models often overlooks basic energy-saving practices [86], and typically results in higher runtime and energy consumption compared to human-written solutions [91].

Emissions and Pollution. Several studies noted the concerns on this dimension and discussed about the solutions [25, 52, 86]. For example, the organization can reduce the carbon emissions by locating the data center in regions with cleaner energy grids or by sourcing electricity from low-carbon or renewable providers [92].

Excluded Dimensions. We exclude some dimensions that are inapplicable to the field of CodeGen: biodiversity conservation under the environment dimension[8, 75], animal ethics (animal welfare) [48, 65], health and safety (in medicine or chemistry) [8, 45], and fair trade (in economy)[46, 82].

Finding 1: Our literature review revealed that ES-CODEGEN are defined by 10 dimensions from 3 areas, including 19 aspects.

³Billy Perrigo (2023). OpenAI Used Kenyan Workers on Less Than \$2 Per Hour To Make ChatGPT Less Toxic. *Time*. Available at: <https://time.com/6247678/openai-chatgpt-kenya-workers/>. Accessed: 2025-05-07.

4 Ethical Considerations

The research ethics committee of our institution requires that each researcher doing research involving human subjects obtain certification on a research ethics course (i.e., an online course with tutorials and exercises on ethics guidance) so all researchers involved have completed the course before getting ethical approval. Then, we submitted the methodology for the study (including the procedures to gather contact information, recruitment materials, survey questions), and received ethical approval from the research ethics committee of the institution directly involved in the survey.

At the start of the survey, we obtained informed consent from participants. We stressed that this is an anonymous survey, participation is voluntary, and they can quit the survey at any time, and incomplete responses will not be recorded. As fair labor is one key dimension in ethical sourcing, we provide five Amazon gift cards valued at 50 (in our designated currency) for every 50 participants as compensation, complying with the minimum wage requirement in our region. To protect the participants' anonymity, we stressed that participation in the lucky draw is optional, and we collected their email addresses to distribute the gift cards in a separate, independent questionnaire.

5 Survey Design and Results Analysis

Building upon the dimensions that we obtained via our literature review, we designed a questionnaire to elicit developers' perspectives on whether these dimensions are appropriate for defining "ethically sourced code generation," and to understand their underlying reasons and concerns. In addition, we investigated the broader significance of ES-CODEGEN, including scenarios in which such ethical considerations are most critical.

5.1 Participants Recruitment

Our survey targeted stakeholders involved in the research and the development of CodeGen models. To identify these stakeholders, we invited participants via several rounds of invitations: (1) we shared the link of our survey through social media posts, (2) we collected a list of emails of authors of the relevant papers in our literature review (we only kept authors relevant to LLM research and ethical issues), leading to a total of 831 emails (out of these emails, we received six replies saying that they would not participate), (3) we contacted researchers within our social network and sent emails to mailing lists for software engineering and AI researchers.

Impacted Users. In the last round of invitations, we included GitHub users who have opted-out from the Stack dataset [19], as they were affected by ethically sourced issues. Our email started with “As you requested to opt-out your repositories from The Stack dataset, we would like to invite...”. As these users have experience of their repository being used as training data and chose to opt-out, their opinions are invaluable (we call them *impacted users*). Interestingly, we observe that the response rate for impacted users is relatively high, i.e., we obtain six new responses in our last round of invitations out of 50 emails sent (12.0% response rate compared to 3.3% for previous invitations), showing that *impacted users are more likely to share their views on ES-CODEGEN*.

As a targeted survey on CodeGen, we listed several inclusion criteria at the start of the survey and in our invitation email: (1) adults

aged 18 and over, (2) have at least one year of programming experience, (3) have at least some familiarities with CodeGen models (as a user, developer, or researcher), (4) not affiliated with the research group conducting the survey (to avoid conflict of interest). Finally, we received 34 survey responses, of which 32 were complete and valid (i.e., we analyzed the results from the 32 ones).

5.2 Survey Design

We designed the survey questionnaires based on our literature review, general guidelines for survey design [36], and SE-specific guidelines [49, 50]. The survey includes five types of questions: (1) short-answer open-ended questions, (2) multiple-choice questions, (3) rating questions (with 5-point Likert scale), (4) matrix table questions, and (5) ranking questions. The survey consists of the six sections below:

1. Background. This consists of five multiple choice questions and three open-ended questions, asking about the participant's non-identifiable background information. Our goal is to check whether participants are familiar with programming, CodeGen models, and the concept of "ethically sourced".

2. Open-ended Questions on Defining ES-CODEGEN. The second section investigated participants' opinions on what should be considered when defining "ethically sourced code generation" and whether they have ever encountered ethical sourcing issues through two open-ended questions. To minimize potential bias and to allow participants to provide the definition in their own words, we posed these open-ended questions prior to presenting our collected dimensions, thereby reducing the likelihood that participants would be influenced by our taxonomy.

3. Dimension Evaluation. Here we first described each dimension and asked about the relevance of each to the definition of ES-CODEGEN through 10 rating questions, with options rated in a five-point Likert scale (from "Not relevant (1)", to "Very relevant(5)"). Secondly, we asked which development stages and artifacts (columns) in the development of CodeGen models they believe each dimension (rows) applies to, with four matrix table questions.

4. Real-World Experiences and Reflections. This part asked about participants' personal reasons for rating the dimensions, their real-world experiences or thoughts related to ES-CODEGEN, and potential impacts of UnES-CODEGEN by four open-ended questions and one multiple choice question.

5. Trade-offs and Acceptable Costs. This section has five multiple-choice questions on respondents' views on the acceptable costs of making a CodeGen model ethically sourced, and one ranking question on the relative importance of each trade-off option.

6. Post-survey Reflections. The final section asked participants to reflect on their experience with this survey. For example, whether they think that ES-CODEGEN is more important than they thought after taking the survey (from "Strongly disagree (1)", to "Strongly agree(5)"), which dimensions were ignored by them before taking this survey (by a multiple choice question), and whether the dimensions are generalizable to other fields beyond CodeGen (e.g., voice generation) (from "Definitely not (1)", to "Definitely yes(5)"). This part includes three five-point Likert scale rating questions, one multiple choice question and one open-ended question.

6 Survey Results

We use Qualtrics to collect survey responses. For the open-ended questions, we use an open coding approach where two authors (coders) performed open coding on all the 32 valid responses. They independently assigned one or more codes to each valid response. To assess inter-rater reliability for these multi-label annotations, we used the approach described in Section 3.3. The results show an overall label-level Cohen's Kappa of 0.93 and an average Jaccard Index of 0.88, indicating a high level of inter-rater agreement.

Background of Survey Participants. Table 3 shows the general demographic information of our survey participants. Most of our participants are researchers (65.6%), have more than ten years of coding experience (50.0%), are knowledgeable (46.9%) about code generation models, and use the models frequently (daily (50.0%) or weekly (28.1%)). Most of them (75.0%) are familiar with more than 3 code generation models, and 90.6% of them have at least one year of experience working with code generation models. Most (40.6%) are somewhat familiar with ethically sourced products, while the same proportion of participants are not familiar with ES-CODEGEN. For the descriptions of different levels of familiarity, please refer to the specific question and options in our full survey.

6.1 RQ1: Various Dimensions that Define ES-CODEGEN

Before asking participants to evaluate dimensions in our taxonomy, we collected the definitions provided by them via an open-ended question, and summarized the following dimensions:

Overlapping Dimensions. Most of the mentioned dimensions overlapped with our dimensions, indicating our collected taxonomy is general, and matches well with the stakeholders' perspective of ES-CODEGEN. Notably, 46.9% mentioned about licensing or copyright; 25.0% mentioned environment; 21.9% mentioned privacy (e.g., "*Prompting data should not be stored, or should be stored securely, and deleted on demand*"), fairness (e.g., "*minimizing bias*"), and source acknowledgment (e.g., some comments stated "*to maintain transparency and traceability, we need to apply the concept of supply chain, like tracking the origin of code (i.e., which public code closely aligns with the generated code)*"); "*The source for generating the code should be recorded and traced*"); and 18.8% mentioned consent.

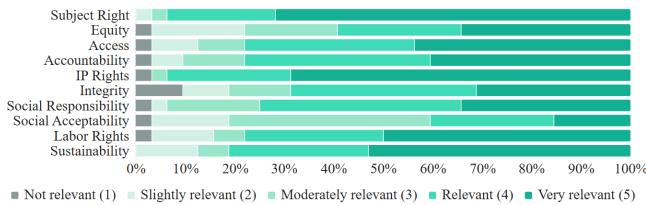
Refined Aspect: Compensation to Authors of Input Training Data. Three out of 32 responses mentioned this so we merged it into "fair wages" under the "labor rights" dimension. As most code generation models mine data from the repository without gathering consent, this indicates *the need to derive a scheme to provide compensation to authors of input training data*.

Refined Aspect by Impacted Users: Opt-in Consent. As impacted users spent effort to submit issues to opt-out from the Stack dataset, we notice that three users suggested about opt-in consent instead of the current opt-out consent policy. Notably, one comment defined ES-CODEGEN as "*Models should only be trained on code voluntarily offered up by developers, on an opt-in basis, and never scraped without consent, including even MIT licensed open source code, which was released as such before the proliferation of LLMs*".

New Dimension: Code Quality. We did not include code quality as a dimension in our initial taxonomy because while metrics such

Table 3: Background of survey participants. Not Familiar (1), Slightly Familiar (2), Somewhat Familiar (3), Knowledgeable (4), Expert (5)

Role	Coding Experience (y)	Familiarity with ES-Products	Familiarity with ES-CODEGEN
Researcher 65.6% (21)	11–20 37.5% (12)	Somewhat familiar 40.6% (13)	Not familiar 40.6% (13)
Software Developer 15.6% (5)	6–10 28.1% (9)	Knowledgeable 28.1% (9)	Slightly familiar 25.0% (8)
Student 12.5% (4)	4–5 18.8% (6)	Slightly familiar 21.9% (7)	Somewhat familiar 18.8% (6)
AI/ML Engineer 6.3% (2)	>21 12.5% (4)	Not familiar 6.3% (2)	Knowledgeable 15.6% (5)
-	2–3 3.1% (1)	Expert 3.1% (1)	Expert 0.0% (0)
Familiarity with CodeGen	# of Familiar CodeGen	Frequency of Using CodeGen	Experience of Working with CodeGen (y)
Knowledgeable 46.9% (15)	4–6 50.0% (16)	Daily 50.0% (16)	2 40.6% (13)
Expert 28.1% (9)	2–3 25.0% (8)	Weekly 28.1% (9)	3 31.3% (10)
Somewhat familiar 18.8% (6)	7–10 15.6% (5)	Monthly 9.4% (3)	1 12.5% (4)
Slightly familiar 6.3% (2)	11–20 6.3% (2)	Yearly 6.3% (2)	0 9.4% (3)
Not familiar 0.0% (0)	21+ 3.1% (1)	Never 6.3% (2)	4 6.3% (2)

**Figure 3: Relevance of Dimensions in Defining ES-CODEGEN**

as code quality and accuracy are important to assess the effectiveness of code generation models, they are typically regarded as technical quality indicators rather than ethical dimensions. However, 18.8% of participants mentioned code quality (e.g., output quality, accuracy of generated code), with one mentioned “*generation of seemingly correct but in fact faulty code, wasting developers’ time in debugging*”(similar to ethical debates on “hypocrite commit” [2] and nonsensical auto-generated commits [90]). Another example directly related to ethical harm is generating source code with a bomb-making tutorial [80] where the auto-generated code can be used to encourage harmful behavior [89]. Thus, we added it in the taxonomy as a new dimension to offer a refreshed technical view of ES-CODEGEN.

Finding 2: Our summarized dimensions cover most of the dimensions in the definitions provided by the survey participants, with refined aspects consent (opt-in) and fair wage (compensation to authors of training data), and a new dimension on code quality (e.g., accuracy). We also observed that impacted users tend to mention opt-in consent when asked to define ES-CODEGEN.

Figure 3 shows the evaluation results on the dimensions from our taxonomy. The top-3 relevant dimensions participants considered are subject rights, intellectual property rights, and environmental sustainability. Overall, nearly all dimensions in our taxonomy are perceived as relevant or very relevant, with the three lowest-ranked dimensions, i.e., social acceptability (40.6%) and equity (59.4%) integrity (68.8%), were still considered relevant or very relevant by many participants, suggesting that the dimensions in our taxonomy are all relevant to the definition of ES-CODEGEN.

Finding 3: All dimensions in our taxonomy are considered relevant or very relevant, except for social acceptability. The most relevant ones are subject rights, intellectual property rights, and environmental sustainability.

When asked about the ignored dimensions before they took this survey, the top five dimensions are social acceptability (56.3%), social responsibility (46.9%), labor rights (46.9%), access (28.1%), and integrity (25.0%). The least ignored dimension is intellectual property rights, which is consistent with their open-ended answers. Besides, 6.3% of participants selected “All of above” while 12.5% selected “None of above”, indicating that the awareness of ethically sourced issues varies among stakeholders.

Finding 4: Before taking the survey, participants tend to ignore dimensions from the social area when defining ES-CODEGEN. However, most thought that these dimensions are at least moderately relevant in defining ES-CODEGEN.

6.2 RQ2: Applicable Stages and Artifacts

We have introduced the dimensions in Section 3.3. In RQ2, we asked participants to map the applicability of each dimension to various stages and artifacts involved in the supply chain of CodeGen (as shown in Figure 1).

Stages. We collected common stages in the development of CodeGen models from prior work [62, 83, 96], and refined them to better examine where ethical risks may arise: 1) data collection, 2) data annotation, 3) data cleaning and preprocessing, 4) model training and fine-tuning, 5) model evaluation, 6) deployment (integrate the trained model into practical environments, enabling it to perform tasks on designated systems either through real-time interaction or scheduled processing), 7) post-deployment (use model in real-world environment and monitoring its outputs and performance during inference). Specifically, we derive 7) post-deployment (Monitoring/Inference) by combining a) inference [96], and b) model monitoring [83]. Table 4 summarizes the applicability of each dimension to the various stages (based on the results of the survey). We can see that participants thought that all stages are generally applicable for all dimensions (on average, 39.4% answered “All of the above” and at least one participant thought that each stage is

applicable for each dimension). Among all stages, data collection is the most frequently selected to be applicable (50.0% on average), with data annotation the second (40.3%).

Artifacts. We also identified a set of artifacts—components within the model development and usage pipeline where ethically sourced issues may manifest from prior work [3, 83]: 1) training data [25, 34], 2) dependencies (i.e., software components including third-party libraries [3]), 3) model metadata (including learned parameters, hyperparameters [25]), 4) documentation [3], 5) prompt (prompt design and template that may have bias or privacy risks [25, 70]), 6) output (i.e., final product where ethical concerns become visible—ranging from biased code [32] to the inadvertent disclosure of private or confidential information, e.g., in the Samsung–ChatGPT incident [28]). Based on prior work [3, 83], we additionally added “prompt” and “output” as new artifacts, as prior studies [25, 32, 70] noted their importance. Table 5 shows the applicability of each dimension to the various artifacts. Generally, the results are similar to those for the stages (each artifact has been noted as applicable by at least one participant). Among all artifacts, training data is the most frequently selected to be applicable (50.9% on average), with documentation the second (33.1%). Meanwhile, 37.5% thought that all artifacts are applicable, indicating that the entire supply chain of code generation model should be considered for ES-CODEGEN.

Finding 5: All stages and artifacts in the supply chain of ES-CODEGEN are applicable to dimensions. This shows the importance of an ethical supply chain in ES-CODEGEN.

6.3 RQ3: Consequences of UnES-CODEGEN

6.3.1 Personal Experience. Before and after rating our provided dimensions, we asked about participants’ personal experience of facing ethical sourcing issues or situations where ES-CODEGEN was needed. Before rating, 56.3% said “no” for personal experience, and two impacted users noted consent, two noted licensing issues. After rating, only 37.5% said “no”, as most issues are covered by our dimensions. However, some practitioners adopt a more practical point of view of code generation where the usefulness of the code generation model is prioritized over other ethical dimensions, as two participants answered “As long as it functions well and passes the test I don’t concern the sources”. Besides, when asked about thoughts on dimensions, one pointed out potential trade-offs between dimensions, as “these dimensions often conflict in practice. For example, you cannot achieve 100% transparency and privacy simultaneously.”

6.3.2 Potential Impact. We listed nine potential impacts collected from our literature review [37, 52, 61, 63, 88, 92, 99], and asked participants to select all that apply. All listed impacts have at least one selection, and the detailed selection percentages are as follows:

For model developers: 1. Lawsuit issues due to intellectual property violations (87.5%), 2. Security risks related to the misuse of personal data or lack of user consent (87.5%), 3. Harmful outputs that reflect bias, toxic language, or offensive stereotypes (78.1%), 4. Environmental impact (78.1%), 5. Reduced user trust and willingness to use the model (71.9%), 6. Negative public reputation (68.8%), 7. Exclusion of certain users or communities due to limited access or high costs (46.9%). For users: 8. Lawsuits due to using copyrighted code generated by models (90.6%), 9. Criticized for plagiarism due

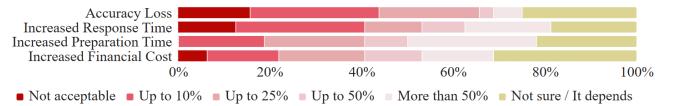


Figure 4: Extent of Acceptable Trade-offs

to using generated code that is too similar to existing code (75.0%). This result is consistent with Finding 3 – most participants cared about intellectual property rights and subject rights.

Open-ended Questions. Before presenting the options, we asked an open-ended question about potential impacts and collected the following new impacts: 1. Exploitation of open-source developers (15.6%, as one comment stated “*human developers not getting the recognition or compensation they deserve*”), 2. Low-quality or unreliable generated code (12.5%, e.g., “*buggy code*”), 3. “*Monopolization of generative AI*” (9.4%), 4. Impact on open-source communities (9.4%, as “*more projects being closed source to avoid scraping*”, “*The tragedy of the commons. Less people will contribute to the commons*”) 5. Reliance on CodeGen models (6.3%), as one noted in ‘any comment’: “*User-cognitive (effects of reliance on code generation models, dependency)*”, 6. Fuel of dishonesty (3.1%), 7. Unemployment (3.1%).

Finding 6: Among the potential impact options, most participants cared about the lawsuit issues due to IP violations on users (90.6%) and model developers (87.5%), then security risks on model developers (87.5%). The top-3 new important impacts noted include 1. Exploitation of developers (15.6%), 2. Low-quality generated code (12.5%), 3. Monopolization of generative AI and impact on open-source communities (9.4%).

6.4 RQ4: Trade-offs & Contamination

Ranking Trade-offs. As noted in studies from our literature review [12, 25], developing CodeGen models needs to balance key trade-offs. Hence, we asked participants to rank the relative importance of five potential trade-offs for ES-CODEGEN, with rank=1 the most important. The result shows that the top-ranked is “accuracy loss” (average rank 1.84), followed by “response time or latency” (2.69), “data contamination” (3.09), “financial or resource cost” (3.28), and “data preparation time” (4.09).

Acceptable Extents. Figure 4 shows the extent to which participants are willing to accept each trade-off (except for contamination), in terms of percentage loss or increase. For accuracy loss (top-ranked previously), 15.6% selected that no loss is acceptable, and 28.1% selected that up to 10% loss is acceptable. For data preparation time (lowest-ranked previously), 28.1% selected that more than 50%, 9.4% selected up to 50%, and no one selected unacceptable.

Acceptable Contamination. We provided six contamination types derived from artifacts and let them select all that apply. As a result, the most acceptable ones are automatically generated content (53.1%) and bot submissions (34.4%), followed by data with unclear or unknown dependencies (28.1%), using closed-source models (28.1%), unknown or unrevealed data sources (25.0%), and data with incompatible license within the dependencies (25.0%). Among all, 31.3% of participants thought no contamination is acceptable.

Table 4: For each dimension, which stages of ES-CODEGEN model development it can apply to?

Dimension	Data Collection	Data Annotation	Data Cleaning	Model Training	Model Evaluation	Model Deployment	Post-deployment	None of Above	All of Above
Subject Right (D1)	65.6%	37.5%	31.3%	37.5%	31.3%	28.1%	25.0%	0.0%	31.3%
Equity (D2)	50.0%	34.4%	34.4%	34.4%	43.8%	31.3%	25.0%	0.0%	37.5%
Access (D3)	40.6%	31.3%	28.1%	28.1%	37.5%	46.9%	37.5%	0.0%	34.4%
Accountability (D4)	46.9%	43.8%	40.6%	40.6%	50.0%	43.8%	31.3%	0.0%	40.6%
Intellectual Property (D5)	65.6%	46.9%	34.4%	40.6%	34.4%	31.3%	25.0%	0.0%	34.4%
Integrity (D6)	56.3%	37.5%	46.9%	34.4%	43.8%	21.9%	25.0%	3.1%	31.3%
Social Responsibility (D8)	50.0%	40.6%	37.5%	34.4%	37.5%	28.1%	28.1%	3.1%	40.6%
Social Acceptability (D9)	50.0%	53.1%	50.0%	28.1%	31.3%	21.9%	25.0%	3.1%	37.5%
Labor Right (D10)	40.6%	50.0%	37.5%	25.0%	25.0%	21.9%	25.0%	6.3%	50.0%
Sustainability (D11)	34.4%	28.1%	21.9%	46.9%	37.5%	37.5%	34.4%	0.0%	56.3%
Average	50.0%	40.3%	36.3%	35.0%	37.2%	31.3%	28.1%	1.6%	39.4%

Table 5: For each dimension, which artifacts of ES-CODEGEN model development it can apply to?

Dimension	Training Data	Dependencies	Model Metadata	Documentation	Prompt	Output	None of Above	All of Above
Subject Right	68.8%	37.5%	28.1%	34.4%	15.6%	15.6%	0.0%	31.3%
Equity	59.4%	15.6%	25.0%	37.5%	25.0%	53.1%	3.1%	21.9%
Access	31.3%	34.4%	37.5%	53.1%	28.1%	37.5%	0.0%	40.6%
Accountability	50.0%	46.9%	37.5%	50.0%	25.0%	31.3%	0.0%	43.8%
Intellectual Property	59.4%	40.6%	31.3%	34.4%	15.6%	15.6%	0.0%	37.5%
Integrity	59.4%	31.3%	9.4%	21.9%	12.5%	21.9%	3.1%	31.3%
Social Responsibility	43.8%	31.3%	12.5%	28.1%	18.8%	28.1%	3.1%	43.8%
Social Acceptability	50.0%	12.5%	18.8%	28.1%	21.9%	37.5%	0.0%	34.4%
Labor Right	46.9%	28.1%	18.8%	21.9%	15.6%	18.8%	3.1%	46.9%
Sustainability	40.6%	31.3%	28.1%	21.9%	18.8%	31.3%	3.1%	43.8%
Average	50.9%	30.9%	24.7%	33.1%	19.7%	29.1%	1.6%	37.5%

Finding 7: The most unacceptable trade-off is accuracy loss (most participants (28.1%) thought that acceptable accuracy loss is up to 10%), while the least is data preparation time. The most unacceptable contamination types are unrevealed data sources and data with incompatible licenses within the dependencies. 31.3% thought that no contamination is acceptable.

Finding 8: Most participants thought that none of the currently available CodeGen models aligns with the definition of ES-CODEGEN or only partially aligns, suggesting a considerable gap between current models and ES-CODEGEN.

6.5 Ethically Source Code Generation Models

After learning about the various dimensions, most participants (68.8%) thought that none of the existing code generation models aligns with the definition of ES-CODEGEN. Meanwhile, 21.9% felt that some tools partially aligned, while two participants answered "yes" and one indicated "no idea". Some practitioners provided examples of partially ethically sourced models, including Llama, Qwen, and Anthropic Claude Code, but these models still need to improve on dimensions like transparency and opt-in consent.

Those who answered "yes" mentioned StarCoder and OLMo2, as one impacted user said "*OLMo2 uses ethically sourced data and doesn't use data without consent while still getting very good performance. They also list all the licenses of which they are using the data under*". Based on the OLMo2's report [72], we notice that as OLMo2 also relied on the Stack dataset, it may share the similar problem of opt-in consent. Among the six impacted users, three answered "no", two noted "partially" (e.g., "*Most tools still lack full transparency and licensing clarity. However, tools that provide citation, allow opt-outs for training data, or offer explainability features are moving in the right direction...*"), and one noted "yes", showing that most impacted users are unsatisfied with the practices of current models.

6.6 Post-Survey Reflections

More than half of the participants agreed (56.3%) or strongly agreed (18.8%) that their understanding of ES-CODEGEN has improved, with the remaining (25.0%) selecting neutral. Similarly, most participants agreed (43.8%) or strongly agreed (15.6%) that this survey made them feel that the issues of ES-CODEGEN were more important than they previously thought. 37.5% selected neutral, and only 3.1% disagreed. Finally, most participants believed that the dimensions and definitions discussed in this survey could be generalized to other fields beyond code generation (e.g., image generation, voice generation). Specifically, 46.9% chose probably yes (4), 43.8% chose definitely yes (5), and 9.4% chose not sure.

Finding 9: Our survey helped improve more than half of the participants' understanding of ES-CODEGEN, raised awareness, and affirmed the generalizability of the definition of ES-CODEGEN.

7 Discussions and Implications

We highlight several implications for the research community:
Multidimensional View in Defining ES-CODEGEN. Our literature review identified 10 dimensions that define ES-CODEGEN (Finding 1), whereas our survey further refined the "labor rights" dimension by stressing the compensation to authors of training data

(Finding 2). Moreover, participants acknowledged that they tend to ignore dimensions in the social area (social acceptability, social responsibility, and labor rights) (Finding 4). As all these dimensions are considered relevant or very relevant in defining ES-CODEGEN (Finding 3), our study indicates *the need to adopt a multidimensional view in defining ES-CODEGEN, and help raise awareness about various dimensions towards ES-CODEGEN* (Finding 9).

Code Quality as a New Technical Dimension of ES-CODEGEN. Stakeholders have expressed the importance of incorporating code quality as a dimension of ES-CODEGEN (Finding 2), and indicated the impact of low quality code being generated (Finding 6). We believe that this *offers a refreshed technical view of ES-CODEGEN, and stresses the importance of adopting technical lens* while considering ethical dimensions of ES-CODEGEN.

Importance of Opt-in Consent Noted by Impacted Users. We observe that impacted users who are directly affected by UnES-CODEGEN are more willing to express their opinions on ES-CODEGEN. Instead of taking initiative to opt-out from the Stack dataset, they also preferred opt-in consent (Finding 2). This shows that *simply enforcing informed consent may be insufficient* as a specific mechanism (opt-in) is preferred over existing practices (opt-out). As data collection for training Code LLMs usually involves mining data from millions of repositories, and these repository owners may not provide their contact information, it is challenging to obtain opt-in consent from these repository owners. Hence, for open-source repository owners, they should consider adding Non-AI licenses to their repositories [4] to explicitly express disagreement with the use of their data for AI training. For AI developers or Data miners, they should *design practical yet ethical approaches to obtain large-scale opt-in consent from developers*, such as developing a consent-management system for open-source repositories.

Consequences of UnES-CODEGEN. Stakeholders noted that there are many potential consequences of UnES-CODEGEN where most participants cared about lawsuit issues that may be faced by users and model developers (Finding 6). Participants also suggested new potential impacts from the perspective of developers and users: 1. Exploitation of developers, 2. Low-quality generated code, 3. Monopolization of generative AI, and 4. Impact on open-source communities. These consequences *help strengthen the motivations of our study and raise awareness of the importance of ES-CODEGEN*.

Ethically Sourced Supply Chain in ES-CODEGEN. As Finding 5 revealed that all stages and artifacts in the supply chain of ES-CODEGEN are considered to be relevant to the dimensions, our study *serves as motivation to ensure that the entire supply chain of code generation is ethically sourced*. Moreover, we also observe that most studies focus on evaluating the generated output [32, 57–59, 99], neglecting other important stages (e.g., data annotation) and artifacts (e.g., documentation). In the future, it is worthwhile to systematically investigate the ethical perspectives of these important stages and artifacts.

Acceptable Trade-offs and Contamination. As most participants noted that code quality is an important dimension, Finding 7 revealed that most noted that accuracy loss is unacceptable (they can only accept up to 10% loss). Some thought that no contamination is acceptable, whereas most thought that they cannot accept contamination from unknown or unrevealed data sources and data with incompatible licenses within the dependencies. As removing

data contamination may result in less data that may in turn lead to accuracy loss, there is *a need to balance the trade-off between accuracy and other dimensions* (e.g., informed consent and IP rights).

Gaps in CodeGen Models. Finding 8 revealed that most participants thought that none of the current CodeGen models aligns with the definition of ES-CODEGEN or only partially aligns. Practitioners noted that current models still need to improve on dimensions like transparency and opt-in consent, indicating the *gaps in existing code generation models*. This calls for the need to design practical techniques to improve CodeGen models on these dimensions.

8 Threats to Validity

External. Our study derives the taxonomy based on our literature review and survey. As ethical issues are trending topics, additional newer literature might have exposed more dimensions in defining ES-CODEGEN. As most dimensions in our taxonomy overlaps with the participants' definitions, our survey results show that our taxonomy are quite general and comprehensive. Our findings may not generalize beyond the survey results provided by the 32 responses. We mitigate this by sending multiple rounds of invitations to obtain diverse options from various stakeholders.

Internal. To reduce researcher bias in open-ended response coding, we use an iterative coding process for all disagreements to reach a consensus. To ensure that we covered stakeholders with diverse backgrounds, we used diverse strategies to invite participants (e.g., posting to mailing lists, contacting stakeholders through our professional network, and emailing impacted users by searching for submitted GitHub issues). However, the low response rate and self-selection bias may have affected the results by attracting participants interested in the survey topic. We designed our survey questions by following best practices and survey design guidelines.

9 Conclusion

This paper introduces the notion of ES-CODEGEN that encompasses 11 different ethical dimensions. To build a broad yet relevant taxonomy of ES-CODEGEN, we perform a two-phase literature review where we reviewed 803 papers to identify 71 relevant papers with 10 initial dimensions of ES-CODEGEN. Based on our literature review, we surveyed 32 practitioners to identified various dimensions, consequences, artifacts, and stages relevant to ES-CODEGEN, leading to 11 dimensions of ES-CODEGEN with new dimension focusing on quality. The evaluation of our survey showed most practitioners either agreed or strongly agreed that our survey help improve their understanding of ES-CODEGEN. As practitioners noted that none of existing code generation models aligns with the definition of ES-CODEGEN, our findings call for the need to design approaches towards ES-CODEGEN.

10 Acknowledgement

We acknowledge the support of the Government of Canada's New Frontiers in Research Fund (NFRF), [NFRFE-2024-00612], the insightful suggestions from reviewers, and detailed answers from all survey participants. We also express our sincere gratitude to Professor Yu Huang and Professor Diego Elias Costa for their invaluable help in this research.

References

- [1] 2024. UNITED STATES DISTRICT COURT SOUTHERN DISTRICT OF NEW YORK. <https://www.courthousenews.com/wp-content/uploads/2024/11/raw-story-openai-copyright-complaint.pdf> Accessed: 2025-05-13.
- [2] 2021. Here's Why University of Minnesota is Getting Banned From Contributing to Linux Kernel Code. <https://news.itsfoss.com/hypocrite-commits/>. Accessed: 2025-09-24.
- [3] 2025. Key components of an AI-BOM. <https://www.wiz.io/academy/ai-bom-ai-bill-of-materials>. Accessed: 2025-07-13.
- [4] 2025. Non-AI Licenses. <https://github.com/non-ai-licenses/non-ai-licenses/>. Accessed: 2025-11-08.
- [5] Victor Ibukun Adebayo, Patience Okpke Paul, and Nsisong Louis Eyo-Udo. 2024. Sustainable procurement practices: Balancing compliance, ethics, and cost-effectiveness. *GSC Advanced Research and Reviews* 20, 1 (2024), 098–107.
- [6] Adobe. 2023. Adobe Unveils Firefly, a Family of new Creative Generative AI. <https://news.adobe.com/news/news-details/2023/Adobe-Unveils-Firefly-a-Family-of-new-Creative-Generative-AI/default.aspx>. Access: 2025-05-07.
- [7] Jawad Alzeer, Khaled Abou Hadeed, Farhan Tufail, et al. 2025. The halal positive list: streamlining the path to certification. *Halalsphere* 5, 1 (2025), 1–7.
- [8] Mahaveer Ameta, Dr. Himani Tiwari, and Dr. Kaushal Kishor Chandrul. 2024. Green Beauty: The Growing demand for Organic Cosmetics. *European Journal of Modern Medicine and Practice* 4, 5 (May 2024), 383–389.
- [9] Upendra Singh Aswal, Pappala Mohan Rao, BR Supreeth, G Manikandan, R Srinivasan, and Mohit Tiwari. 2024. Empowering Human-Machine Communication in Code Generation with NLP. In *2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCT)*. IEEE, 138–142.
- [10] Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, et al. 2021. Program synthesis with large language models. *arXiv preprint arXiv:2108.07732* (2021).
- [11] Nor Azrina Nadiha Azmi, Najihah Mohd Noor, Amal Elgharbawy, and Noviyani Darmawan. 2024. Combining science and faith for better quality and safety in halal toyibyan products. *Halal Studies and Society* 1, 3 (2024), 12–14.
- [12] Solon Barocas, Anhong Guo, Ece Kamar, Jacquelyn Krones, Meredith Ringel Morris, Jennifer Wortman Vaughan, W Duncan Wadsworth, and Hanna Wallach. 2021. Designing disaggregated evaluations of ai systems: Choices, considerations, and tradeoffs. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 368–378.
- [13] Stephanie Barrientos. 2012. *Ethical sourcing in the global food system*. Routledge.
- [14] Brett A Becker, Paul Denny, James Finnie-Ansley, Andrew Luxton-Reilly, James Prather, and Eddie Antonio Santos. 2023. Programming is hard—or at least it used to be: Educational opportunities and challenges of ai code generation. In *Proceedings of the 54th SIGCSE TS V. 1*. 500–506.
- [15] Jean-Christophe Bélisle-Pipon, Maria Powell, Renee English, Marie-Françoise Malo, Vardit Ravitsky, Bridge2AI-Voice Consortium, and Yael Bensoussan. 2024. Stakeholder perspectives on ethical and trustworthy voice AI in health care. *Digital Health* 10 (2024), 20552076241260407.
- [16] Jean-Christophe Bélisle-Pipon and Vardit Ravitsky. 2025. Consent Is Dead, Long Live Ethical Oversight: Integrating Ethically Sourced Data into Demonstrated Consent Models. *The American Journal of Bioethics* 25, 4 (2025), 112–115.
- [17] Jean-Christophe Bélisle-Pipon, Vardit Ravitsky, Bridge2AI-Voice Consortium, and Yael Bensoussan. 2023. Individuals and (synthetic) data points: using value-sensitive design to foster ethical deliberations on epistemic transitions. *The American Journal of Bioethics* 23, 9 (2023), 69–72.
- [18] BigCode Project. 2024. Discussion about opt-out. <https://github.com/bigcode-project/opt-out-v2/issues/2091>. Accessed: 2025-07-08.
- [19] BigCode Project. 2024. Opt-out requests v2. <https://github.com/bigcode-project/opt-out-v2>. Accessed: 2025-07-08.
- [20] Brendon K Billings, Beverley Kramer, Tanya N Augustine, Desiré Brits, Erin F Hutchinson, Elena Libhaber, and Goran Štrkalj. 2024. Leading the transition to ethical human body sourcing in Africa: The South African experience. *Annals of Anatomy-Anatomischer Anzeiger* 254 (2024), 152263.
- [21] Bloomberg. 2024. Adobe's 'Ethical' Firefly AI Was Trained on Midjourney Images. <https://www.bloomberg.com/news/articles/2024-04-12/adobe-s-ai-firefly-used-ai-generated-images-from-rivals-for-training>.
- [22] Mick Blowfield. 2000. Ethical sourcing: a contribution to sustainability or a diversion? *Sustainable Development* 8, 4 (2000), 191–200.
- [23] Roja Boina and Alekhya Achanta. 2023. Balancing language brilliance with user privacy: A call for ethical data handling in ChatGPT. *International Journal of Science and Research* 10, 6 (2023), 657–711.
- [24] Matthew Butterick and Joseph Saveri Law Firm. 2022. GitHub Copilot Investigation. <https://githubcopilotinvestigation.com/> Accessed: 2025-05-31.
- [25] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2023. Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research* 24, 240 (2023), 1–113.
- [26] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46.
- [27] Daniela S Cruzes and Tore Dyba. 2011. Recommended steps for thematic synthesis in software engineering. In *ESEM*. IEEE, 275–284.
- [28] Dark Reading Staff. 2023. Samsung Engineers Leak Sensitive Data to ChatGPT, Prompting Warnings on AI Use in the Workplace. <https://www.darkreading.com/vulnerabilities-threats/samsung-engineers-sensitive-data-chatgpt-warnings-ai-use-workplace> Accessed: 2025-05-07.
- [29] Jeffrey Dustin and Anna Tong. 2023. Google, one of AI's biggest backers, warns own staff about chatbots. (2023). <https://www.reuters.com/technology/google-one-ais-biggest-backers-warns-own-staff-about-chatbots-2023-06-15/> Accessed: 2025-05-13.
- [30] Kayla DePalma, Izabel Miminoshvili, Chiara Henselder, Kate Moss, and Eman Abdullah AlOmar. 2024. Exploring ChatGPT's code refactoring capabilities: An empirical study. *Expert Systems with Applications* 249 (2024), 123602.
- [31] John Elkington. 1998. Partnerships from cannibals with forks: The triple bottom line of 21st-century business. *Environmental quality management* 8, 1 (1998), 37–51.
- [32] Yunhe Feng, Sreecharan Vanam, Manasa Cherukupally, Weijian Zheng, Meikang Qiu, and Haihua Chen. 2023. Investigating code generation performance of ChatGPT with crowdsourcing social data. In *47th COMPSAC*. IEEE, 876–885.
- [33] Dessislava S Fessenko. 2024. Ethical Requirements for Achieving Fairness in Radiology Machine Learning: An Intersectionality and Social Embeddedness Approach. *Journal of Health Ethics* 20, 1 (2024), 4.
- [34] Manuel B Garcia. 2025. Teaching and learning computer programming using ChatGPT: A rapid review of literature amid the rise of generative AI technologies. *Education and Information Technologies* (2025), 1–25.
- [35] Philipp Goebel, Carsten Reuter, Richard Pibernik, and Christina Sichtmann. 2012. The influence of ethical culture on supplier selection in the context of sustainable sourcing. *International Journal of Production Economics* 140, 1 (2012), 7–17.
- [36] Robert M Groves, Floyd J Fowler Jr, Mick P Couper, James M Lepkowski, Eleanor Singer, and Roger Tourangeau. 2011. *Survey methodology*. John Wiley & Sons.
- [37] Maanak Gupta, CharanKumar Akiri, Kshitiz Aryal, Eli Parker, and Lopamudra Praharaj. 2023. From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access* 11 (2023), 80218–80245.
- [38] Vikram Gupta, Rini Sharon, Ramit Sawhney, and Debdoott Mukherjee. 2022. Adiman: Abuse detection in multilingual audio. In *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 6172–6176.
- [39] Thilo Hagendorff. 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and machines* 30, 1 (2020), 99–120.
- [40] Lina Hamad, Choong Key Chekar, Chloe Anthias, and Laura Machin. 2024. Facilitating the ethical sourcing of donor hematopoietic stem cells for cell and gene therapy research and development. *Regenerative Medicine* 19, 6 (2024), 317–326.
- [41] Catarina Holmsten-Carrizo and Cecilia Mark-Herbert. 2017. Luxury marketing strategies related to ethical sourcing. In *Sustainable Luxury*. Routledge, 43–54.
- [42] Samireh Jalali and Claes Wohlin. 2012. Systematic literature studies: database searches vs. backward snowballing. In *Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement*. 29–38.
- [43] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature machine intelligence* 1, 9 (2019), 389–399.
- [44] Peter Kamstra, Jane Farmer, Anthony McCosker, Fergus Gardiner, Hazel Dalton, David Perkins, Luis Salvador-Carulla, and Nasser Bagheri. 2023. A novel mixed methods approach for integrating not-for-profit service data via qualitative geographic information system to explore authentic experiences of ill-health: a case study of rural mental health. *Journal of Mixed Methods Research* 17, 4 (2023), 419–442.
- [45] Jung-Hyun Kim, Hye-Yeong Jo, Hye-Yeong Ha, and Yong-Ou Kim. 2021. Korea national stem cell bank. *Stem cell research* 53 (2021), 102270.
- [46] Seongtae Kim and Sangho Chae. 2022. Shareholder value effects of ethical sourcing: comparing reactive and proactive initiatives. *Journal of Business Ethics* 179, 3 (2022), 887–906.
- [47] Seongtae Kim, Claudia Colicchia, and David Menachof. 2018. Ethical sourcing: An analysis of the literature and implications for future research. *Journal of Business Ethics* 152 (2018), 1033–1052.
- [48] PVS Kishore, PSS Anusha, DVVN Durga Prasad, K Archana, MPS Tomar, and NKB Raju. 2021. Comparative studies on different embalming fluids in ethically sourced canine cadavers. (2021).
- [49] Barbara A Kitchenham and Shari Lawrence Pfleeger. 2002. Principles of survey research part 2: designing a survey. *ACM SIGSOFT Software Engineering Notes* 27, 1 (2002), 18–20.
- [50] Barbara A Kitchenham and Shari Lawrence Pfleeger. 2002. Principles of survey research: part 3: constructing a survey instrument. *ACM SIGSOFT Software Engineering Notes* 27, 2 (2002), 20–24.
- [51] Jan H Klemmer, Stefan Albert Horstmann, Nikhil Patnaik, Cordelia Ludden, Cordell Burton Jr, Carson Powers, Fabio Massacci, Akond Rahman, et al. 2024. Using ai assistants in software development: A qualitative study on security practices and concerns. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*. 2726–2740.

- [52] Sam Lau and Philip Guo. 2023. From "Ban it till we understand it" to "Resistance is futile": How university programming instructors plan to adapt as more students use AI code generation and explanation tools such as ChatGPT and GitHub Copilot. In *Proceedings of the 2023 ACM Conference on International Computing Education Research–Volume 1*. 106–121.
- [53] Claire Le Goues, Michael Pradel, and Abhik Roychoudhury. 2019. Automated program repair. *Commun. ACM* 62, 12 (2019), 56–65.
- [54] Ke Li, Sheng Hong, Cai Fu, Yunhe Zhang, and Ming Liu. 2023. Discriminating human-authored from chatgpt-generated code via discernable feature analysis. In *IEEE 34th ISSREW*. IEEE, 120–127.
- [55] Raymond Li, Louisa Ben Allal, Yangtian Zi, Niklas Muenninghoff, Denis Kocetkov, Chenghao Mou, Marc Marone, Christopher Akiki, Jia Li, Jenny Chim, et al. 2023. Starcoder: may the source be with you! *arXiv preprint arXiv:2305.06161* (2023).
- [56] Raymond Li, Yangtian Zi, Niklas Muenninghoff, Denis Kocetkov, Chenghao Mou, Marc Marone, Christopher Akiki, LI Jia, Jenny Chim, et al. 2023. StarCoder: may the source be with you! *Transactions on Machine Learning Research* (2023).
- [57] Lin Ling. 2024. Evaluating Social Bias in Code Generation Models. In *Companion Proceedings of the 32nd ACM International Conference on the Foundations of Software Engineering*. 695–697.
- [58] Jianguyu Liu and Siran Li. 2024. Toward artificial intelligence-human paired programming: A review of the educational applications and research on artificial intelligence code-generation tools. *Journal of Educational Computing Research* 62, 5 (2024), 1385–1415.
- [59] Yan Liu, Xiaokang Chen, Yan Gao, Zhe Su, Fengji Zhang, Daoguang Zan, Jian-Guang Lou, Pin-Yu Chen, and Tsung-Yi Ho. 2023. Uncovering and quantifying social biases in code generation. *Advances in Neural Information Processing Systems* 36 (2023), 2368–2380.
- [60] Michael R. Lyu, Baishakhi Ray, Abhik Roychoudhury, Shin Hwei Tan, and Patanamon Thongtanunam. 2025. Automatic Programming: Large Language Models and Beyond. *ACM Trans. Softw. Eng. Methodol.* 34, 5, Article 140 (May 2025), 33 pages. doi:10.1145/3708519
- [61] Wanlun Ma, Yiliao Song, Minhui Xue, Sheng Wen, and Yang Xiang. 2024. The "Code" of Ethics: A Holistic Audit of AI Code Generators. *IEEE Transactions on Dependable and Secure Computing* (2024).
- [62] Abbas Mahmoudabadi and Razieh Akbari Lalaei. 2024. Promoting Project Outcomes: A Development Approach to Generative AI and LLM-Based Software Applications' Deployment. Available at SSRN 4907319 (2024).
- [63] Abdul Majeed and Seong Oun Hwang. 2024. Reliability Issues of LLMs: ChatGPT a Case Study. *IEEE Reliability Magazine* (2024).
- [64] Jet Mboga. 2017. Ethical sourcing to ensure sustainability in the food and beverage industry and eliciting millennial perspectives. *European Journal of Economic and Financial Research* (2017).
- [65] Kaitlin Mitchell and LF Carver. 2024. Investigating the Use of Companion Animal Donor Tissue Banks. *One Health Innovation* 2, 1 (2024).
- [66] Martin Monperrus. 2018. Automatic software repair: A bibliography. *ACM Computing Surveys (CSUR)* 51, 1 (2018), 1–24.
- [67] Veera Harish Muthazhagu and B Surendiran. 2024. Exploring the Role of AI in Web Design and Development: A Voyage through Automated Code Generation. In *ITCCE*. IEEE, 1–8.
- [68] Gerard M Nadal. 2021. Charting a Course toward a New Vaccine Strategy. *The Linacre Quarterly* 88, 3 (2021), 247–253.
- [69] Claudia Negri-Ribalta, Rémi Géraud-Stewart, Anastasia Sergeeva, and Gabriele Lenzi. 2024. A systematic literature review on the impact of AI models on the security of code generation. *Frontiers in Big Data* 7 (2024), 1386720.
- [70] Liang Niu, Shujaat Mirza, Zayd Maradni, and Christina Pöpper. 2023. {CodexLeaks}: Privacy leaks from code generation language models in {GitHub} copilot. In *32nd USENIX Security Symposium*. 2133–2150.
- [71] Kassi Nzalasse, Rishav Raj, Eli Laird, and Corey Clark. 2024. SIG: A Synthetic Identity Generation Pipeline for Generating Evaluation Datasets for Face Recognition. In *International Conference on Pattern Recognition*. Springer, 299–313.
- [72] Team OLMo, Pete Walsh, Luca Soldaini, Dirk Groeneweld, Kyle Lo, Shane Arora, Akshita Bhagia, Yuling Gu, Shengyi Huang, Matt Jordan, et al. 2024. 2 OLMo 2 Furious. *arXiv preprint arXiv:2501.00656* (2024).
- [73] Danielle M Pacia, Vardit Ravitsky, Jan N Hansen, Emma Lundberg, Wade Schulz, and Jean-Christophe Belisile-Pipon. 2024. Early AI lifecycle co-reasoning: ethics through integrated and diverse team science. *The American Journal of Bioethics* 24, 9 (2024), 86–88.
- [74] Nishant A Parikh. 2023. Empowering business transformation: The positive impact and ethical considerations of generative AI in software product management—a systematic literature review. *Transformational Interventions for Business, Technology, and Healthcare* (2023), 269–293.
- [75] Mariano Peluso. 2023. Navigating the coffee business landscape: challenges and adaptation strategies in a changing world. In *Proceedings*, Vol. 89. MDPI, 22.
- [76] James Prather, Paul Denny, Juho Leinonen, Brett A Becker, Ibrahim Albluwi, Michelle Craig, et al. 2023. The robots are here: Navigating the generative ai revolution in computing education. In *Proceedings of the working group reports on innovation and technology in computer science education*. 108–159.
- [77] BigCode Project. 2024. The Stack v2 Dedup. <https://huggingface.co/datasets/bigcode/the-stack-v2-dedup>. Accessed: 2025-05-13.
- [78] Remi Pulverail and F Givaudan. 2013. Ethical Sourcing of Raw Materials. *Sustainability: How the Cosmetics Industry is Greening Up* (2013), 69–95.
- [79] Anaïs Rameau, Satrajit Ghosh, Alexandros Sigaras, Olivier Elemento, Jean-Christophe Belisile-Pipon, Vardit Ravitsky, Maria Powell, et al. 2024. Developing Multi-Disorder Voice Protocols: A team science approach involving clinical expertise, bioethics, standards, and DEI. In *Proc. Interspeech 2024*. 1445–1449.
- [80] Qibing Ren, Chang Gao, Jing Shao, Junchi Yan, Xin Tan, Wai Lam, and Lizhuang Ma. 2024. Codeattack: Revealing safety generalization challenges of large language models via code completion. *arXiv preprint arXiv:2403.07865* (2024).
- [81] Konstantinos I Roumeliotis, Nikolaos D Tselikas, and Dimitrios K Nasiopoulos. 2023. Unveiling Sustainability in ecommerce: GPT-Powered software for identifying sustainable product features. *Sustainability* 15, 15 (2023), 12015.
- [82] Darshak Sanghavi. 2022. Sustainability and Ethical Considerations in Product Management. *Journal of Advances in Developmental Research* 13, 2 (2022).
- [83] Marius Schlegel and Kai-Uwe Sattler. 2023. Management of machine learning lifecycle artifacts: A survey. *ACM SIGMOD Record* 51, 4 (2023), 18–35.
- [84] Roy Schwartz, Jesse Dodge, Noah A Smith, and Oren Etzioni. 2020. Green ai. *Commun. ACM* 63, 12 (2020), 54–63.
- [85] Kristen M Scott, Simone Ashby, David A Braude, and Matthew P Aylett. 2019. Who owns your voice? Ethically sourced voices for non-commercial TTS applications. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. 1–3.
- [86] Samarth Sikand, Rohit Mehra, Vibhu Saujanya Sharma, Vikrant Kaulgud, Sanjay Podder, and Adam P Burden. 2024. Do Generative AI Tools Ensure Green Code? An Investigative Study. In *Proceedings of the 2nd International Workshop on Responsible AI Engineering*. 52–55.
- [87] Steven E Stemler. 2004. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation* 9, 1 (2004).
- [88] Constantine Stephanidis, Gavriel Salvendy, Margherita Antona, Vincent G Duffy, Qin Gao, Waldemar Karwowski, Shin'ichi Konomi, Fiona Nah, Stavroula Ntoua, Pei-Luen Patrick Rau, et al. 2025. Seven HCI Grand Challenges Revisited: Five-Year Progress. *IJHCI* (2025), 1–49.
- [89] Honghao Tan, Haibo Wang, Diany Pressato, Yisen Xu, and Shin Hwei Tan. 2025. Coverage-Based Harmfulness Testing for LLM Code Transformation. *Proceedings of the 40th IEEE/ACM International Conference on Automated Software Engineering, ASE 2025*.
- [90] Linus Torvalds. 2025. Can we please stop this automated idiocy? <https://lore.kernel.org/all/CAHk-wjamixjqNwrr4+UEAwitMOD6Y8-9p4oUZdcjr7fsayQ@mail.gmail.com/> Accessed: 2025-09-07.
- [91] Jonas F Tuttle, Dayuan Chen, Amina Nasrin, Noe Soto, and Ziliang Zong. 2024. Can LLMs Generate Green Code-A Comprehensive Study Through LeetCode. In *15th International Green and Sustainable Computing Conference*. IEEE, 39–44.
- [92] Tina Vartziotis, Ippolyti Dellatolas, George Dasoulas, Maximilian Schmidt, Florian Schneider, Tim Hoffmann, Sotirios Kotsopoulos, and Michael Keckeisen. 2024. Learn to code sustainably: An empirical study on green code generation. In *Proceedings of the 1st LLM4Code*. 30–37.
- [93] James Vincent. 2023. Apple restricts employees from using ChatGPT over fear of data leaks. <https://www.theverge.com/2023/5/19/23729619/apple-bans-chatgpt-openai-fears-data-leak>. Accessed: 2025-05-13.
- [94] Shenao Wang, Yanjie Zhao, Xinyi Hou, and Haoyu Wang. 2025. Large language model supply chain: A research agenda. *ACM Transactions on Software Engineering and Methodology* 34, 5 (2025), 1–46.
- [95] Yanchen Wang and Lisa Singh. 2024. It Is Time to Develop an Auditing Framework to Promote Value Aware Chatbots. *arXiv preprint arXiv:2409.01539* (2024).
- [96] Carole-Jean Wu, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newshe Ardalani, Kiwan Maeng, Gloria Chang, Fiona Aga, Jinshi Huang, Charles Bai, et al. 2022. Sustainable ai: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems* 4 (2022), 795–813.
- [97] Boming Xia, Tingting Bi, Zhenchang Xing, Qinghua Lu, and Liming Zhu. 2023. An empirical study on software bill of materials: Where we stand and the road ahead. In *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, 2630–2642.
- [98] Boming Xia, Dawen Zhang, Yue Liu, Qinghua Lu, Zhenchang Xing, and Liming Zhu. 2024. Trust in software supply chains: Blockchain-enabled sbom and the aibom future. In *Proceedings of the 2024 ACM/IEEE 4th International Workshop on Engineering and Cybersecurity of Critical Systems (EnCyCriS) and 2024 IEEE/ACM Second International Workshop on Software Vulnerability*. 12–19.
- [99] Weixiang Yan and Yuanchun Li. 2022. WhyGen: explaining ML-powered code generation by referring to training examples. In *Proceedings of the ACM/IEEE 44th International Conference on Software Engineering: Companion Proceedings*. 237–241.
- [100] Simon Zadek. 2007. The path to corporate responsibility. In *Corporate ethics and corporate governance*. Springer, 159–172.