

USE OF “SOURCE CODE” AND “AI” IN POLICY DOCUMENTS AND STUDIES

Overview of Policy Documents and Current Research (as of April 2021)

The current EU proposal on the non-disclosure of source code of software in the context of the Joint Statement Initiative on electronic commerce (JSI) is not openly referencing the words “algorithms” or “artificial intelligence (AI)”. However, recent analysis highlights that a legal interpretation of the term “source code of software” would effectively include algorithms and AI as well.¹ One reason for this is a lack of definition provided by the respective parties what a “source code of software” is. Any treaty interpretation in case of disputes over the nature of the term “source code of software” would also take note of the common use of the term. Which is why this note provides an overview below displaying where current policy documents and research already refer to the interchangeable nature of the terms “algorithms”, “AI” and “source code”. The mere fact that other constituencies have mentioned “algorithms” explicitly in their trade agreements along with “source code” does not seem solid enough to argue that “source code” in turns excludes algorithms.

POLICY DOCUMENTS

EU’s Draft AI Legislation

Annex IV on Technical Documentation refers to the descriptions of the main elements of an AI system and the process of its development.

- ❖ Here it also mentions: “the programming code(s), the description of system architecture explaining how software components build on or feed into each other and integrate into the overall processing, the computational resources used to build, test and validate the AI systems;” (Annex IV, b.iii)

OECD: Recommendation of the Council on Artificial Intelligence²

The OECD Recommendation provides a set of internationally agreed principles and recommendations to promote an AI-powered crisis response that is trustworthy and respects human-centred and democratic values.

- ❖ The Recommendation refers to the need for a better AI knowledge including “skills and resources, such as data, code, algorithms, models, research, know-how, training programmes, governance, processes and best practices, required to understand and participate in the AI system lifecycle.”

¹ Kristina Irion (2021): AI Regulation in the European Union and Trade Law, study commissioned by the Federation of German Consumer Organisations (vzbv), on source code interpretation see from p.55.

² <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

Council of Europe: Ad-Hoc Committee on Artificial Intelligence (feasibility study)³

The Council of Europe's CAHAI committee (Ad-hoc Committee on AI) has been instructed to examine the feasibility and potential elements, based on broad multi-stakeholder consultations, of a legal framework for the development, design and application of artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law.

- ❖ With regard to the question of transparency, CAHAI notes that most often it was not clear whether transparency "should be achieved through publishing source code, rendering the algorithmic training data accessible or auditable (while considering the applicable data protection laws) or through some other means. Resolving the challenge of applying these principles in practice and considering potential interdependencies and trade-offs with other desirable properties was hence considered an important issue to be addressed by policy makers." (p. 20, para 74)
- ❖ In enforcing compliance with existing or future rules on AI, CAHAI advises a broad package of initiatives to drive change. This also includes "to drive the widespread adoption of norms such as open access to source code" (p. 55, para 170)

Council of Europe - Towards Regulation of AI Systems, Compilations of Contributions⁴

The CAHAI's Secretariat publication aims to support the ongoing reflections on the analysis of the challenges arising from AI systems and possible regulatory responses. It showcases global perspectives on the development of a legal framework on Artificial Intelligence systems.

- ❖ On the question of AI transparency, the report notes that "most guidelines agree that AI should be transparent to avoid potential problems. But it is not clear whether transparency should be achieved through publishing source code, the underlying databases or some other means." (p.39)
- ❖ With respect to judicial systems and the potential use of AI in judicial decisions, the report notes that "compared to human decisions, and more specifically judicial decisions – the logic behind AI systems does not resemble legal reasoning. Instead they simply execute codes based on a data-centric and mathematical/statistical approach." (p.84)

US Executive Order 13859 of February 11, 2019 Maintaining American Leadership in Artificial Intelligence⁵

The Executive Order establishes an "Enterprise Data and Source Code Inventory" to support discovery and usability in AI Research and Development.

³ <https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da>

⁴ <https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a>

⁵ <https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence>; https://code.gov/assets/data/ai_inventory-guidance.pdf

- ❖ The AI Executive Order references source code as a mechanism for providing data and models. Code.gov allows for discoverability of federal source code by searching AI related keywords. To increase discovery for Federal AI R&D models, agencies shall inform their various offices/department heads associated with R&D that models under development (or procurement) should be captured in Source Code Inventories.
- ❖ Agencies shall include the keyword of "usgartificial-intelligence" for all source code determined to support AI R&D.

RESEARCH / TECH

Automatic Algorithm Specification to Source Code Translation

"In this paper, we describe a translation program that can create a piece of executable code, given the code's algorithmic specification. This program allows the user to specify his/her code using an easy-to-understand, simple-to-write and more or less immutable pseudo code specification. The program will then check the pseudo code for errors, and convert it to a specified language (be it C, Java, or any other language). The program may easily be extended to accommodate different languages. Our program allows the user to focus on just the algorithm, and not on issues related to implementation."

<http://www.ijcse.com/docs/IJCSE11-02-02-51.pdf>

Automatic Algorithm Recognition of Source-Code Using Machine Learning

"As codebases for software projects get larger, reaching ranges of millions of lines of code, the need for computer-aided program comprehension grows. We define one of the tasks of program comprehension to be algorithm recognition: given a piece of source-code from a file, identify the algorithm this code is implementing, such as brute-force or dynamic programming. Most research in this area is making use of pattern matching, which involves much human effort and is of questionable accuracy when the structure and semantics of programs change. Thus, this paper proposes to let go of defined patterns, and make use of simpler features, such as counts of variables and counts of different constructs to recognize algorithms. We then feed these features to a classification algorithm to predict the class or type of algorithm used in this source code. We show through experimental results that our proposed method achieves a good improvement over baseline."

<https://ieeexplore.ieee.org/document/8260630>

Source Code / ML in research papers

"Having source code to go along with a research paper helps a lot in verifying the validity of a machine learning technique and building on top of it. But this is not a requirement for machine learning conferences. (...) But ContributionSecure14 also acknowledges that there are sometimes legitimate reasons for machine learning researchers not to release their code. For example, some authors may train their models on internal infrastructure or use large internal datasets for pretraining. In such cases, the researchers are not at the liberty to publish the code or data along

with their paper because of company policy. "If the authors publish a paper without code due to such circumstances, I personally believe that they have the academic responsibility to work closely with other researchers trying to reproduce their paper", ContributionSecure14 says."

<https://bdtechtalks.com/2021/03/01/papers-without-code-machine-learning-reproducibility/>

New AI model (natural language processing)

"Countless are the business applications that this NLP model named GPT-3, which uses deep learning to produce human-like text, can power – and many have understood that. But, while GPT-3 has great potential for generative value, it is set to fundamentally reshape the dynamics of the AI world and the competitive landscape in the Software-as-a-Service (SaaS) startup space because of its very own business model.

For one, GPT-3 breaks the mold of past AI models, which have traditionally been open source. This gave developers an inside view into the workings of the model and allowed them to add to it. Now with GPT-3, OpenAI provides the ready-made model as a commercial product in the form of a "text in, text out" interface. As such, GPT-3 resets the rules of the AI model game because it does not give away its code but merely offers an easy-to-use application programming interface (API) on a commercial basis, allowing developers to tap into the GPT-3 power while not giving away a peek into its inner workings."

<https://www.forbes.com/sites/hannahmayer/2021/01/02/revolutionary-nlp-model-gpt-3-poised-to-redefine-ai-and-next-generation-of-startups/?sh=6b7ec18e77b3>

Contact

*Verbraucherzentrale
Bundesverband e.V.*

Brussels Office

*Rudi-Dutschke-Str. 17
10969 Berlin*

Buero-bruessel@vzbv.de