

SURVEY STUDY – CODEBOOK

“The Good, the Bad, and the (Un)Usable: A Rapid Literature Review on Privacy as Code”

THEME	CODE	DESCRIPTION
SCOPE	Analyze	The tool helps identifying privacy threats in the source code
	Generate	The tool generates privacy-friendly implementations
TECHNIQUE	Static Taint Analysis	Analysis of the code without executing it. Based in the identification of sinks and sources of personal information. They spot program execution paths in which sensitive data reaches a leakage point without proper sanitization.
	Named Entity Recognition	Natural language processing (NLP) method that identifies predefined categories of objects in a body of text.
	Machine Learning	Statistical algorithms that can learn from data (e.g., K-Anonymity and decision trees).
ABSTRACTION	Abstract Syntax Trees	Data structures that represent and organize a code's syntax (e.g., statements, expressions, and control structures) into the nodes of a tree.
	Call Graphs	Data structures that show how the processes of a program call each other, forming a directed graph.
	Code Property Graphs	Combines various graph structures (e.g., call graphs and abstract syntax trees), capturing the syntactic structure, control flow, and data dependencies of a program.
	Data Property Graphs	Describes the operations existing in a piece of source code and the data types they can access.
	Data Flow Graphs	Tracks how data is passed and modified throughout the program, helping identify where variables are defined, used, and passed.
DESIGN STRATEGIES	Minimize	Collect the least amount of personal data necessary to achieve a certain goal.
	Abstract	Use general data instead of specific details where possible
	Hide	Protect data from unauthorized access through encryption, anonymization, or pseudonymization.
THREATS	Data disclosure	Unauthorized access to personal or sensitive data
	Non compliance	Failing to meet legal or regulatory requirements concerning data privacy
	Unawareness	Not being unaware of how data is collected, processed, or shared.
LIMITATIONS	Manual effort	The approach requires significant manual intervention from the users (e.g., manual annotations in code).
	Performance	The approach's precision is rather low
	Generalizability	The approach has not been tested thoroughly (e.g., across different domains or case studies).
	Scalability	It is difficult to apply the approach in large-scale systems
EVALUATION	Proof of Concept	Preliminary demonstration showing that the tool is viable for a specific purpose.
	Ad-Hoc dataset	A dataset specially curated to assess the performance of that specific tool
	Tool benchmark	A comparison of the tool's performance against others.
	Case study	Application of the tool to a real-world example
	Usability study	Evaluation of the tool with software practitioners (e.g., regarding its ease of use)