From Copilot to Commit: Handling AI-Generated Code in Open Source Crypto Libraries

Dr. Lina Böcker

October 14th, 2026

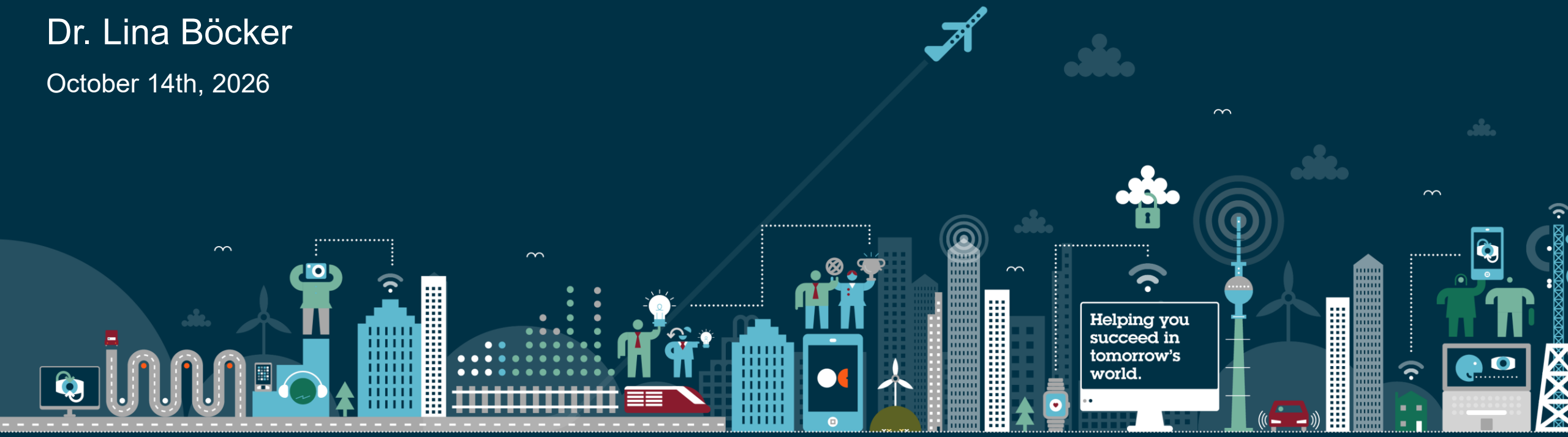
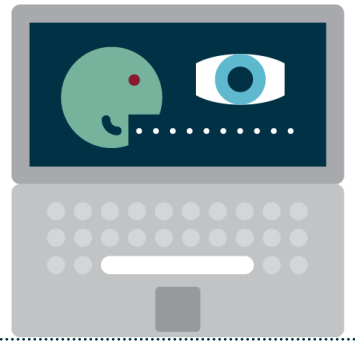


Table of Contents

1 Use of AI Tools for Software Development

2 Use of Open-Source-AI

3 Common Legal Pitfalls in the Use of Open Source Software and Open-Source-AI



1 Use of AI Tools for Software Development

Copyright framework of Open Source

Software and thus also **OSS are protected under copyright law**
→ exploitation rights remain with the author

But: the **author grants third parties rights of use**
→ often subject to the condition of compliance with the license terms

!

Open Source ≠ Free to use without restrictions

The specific type of license dictates such restrictions in form of usage, modification and distribution rules
- **Especially Copyleft-Licenses**

Restrictions of each license and their compatibility with each other must be observed

Current and future risks

AI-generated code is not a future risk. It is already a legal and operational reality today. 42% of developers currently state that AI writes at least half of their code.**

Security

AI models like "Mythos" by Anthropic can systematically detect and exploit vulnerabilities in program code – the same **technology that writes code can also attack it**

Security vulnerability:
In specific repositories, rates of up to 88% Security pass rate: 56%**

Law & Compliance

Risk of **copyright infringement**

EU AI Act: Documentation and governance obligations for AI

Control and Responsibility

If AI-generated code causes errors, the **chain of liability can be unclear**

Supply chains manipulated by AI: Attackers smuggle malicious code through AI-generated open-source contributions

* https://venturebeat.com/data/the-share-of-developers-using-ai-to-write-half-or-more-code-jumped-from-12-to-42-yoy-in-latest-bairesdev-survey
** https://www.veracode.com/resources/analyst-reports/2026-genai-code-security-report/

AI in everyday development – What are we already doing? (1)

Code generation & auto-completion

- suggests code lines while typing
- tackles assigned tasks end-to-end (agents)

Debugging

- analyses complex error messages and stack traces
- suggests likely root causes and concrete fixes

Documentation & explanations

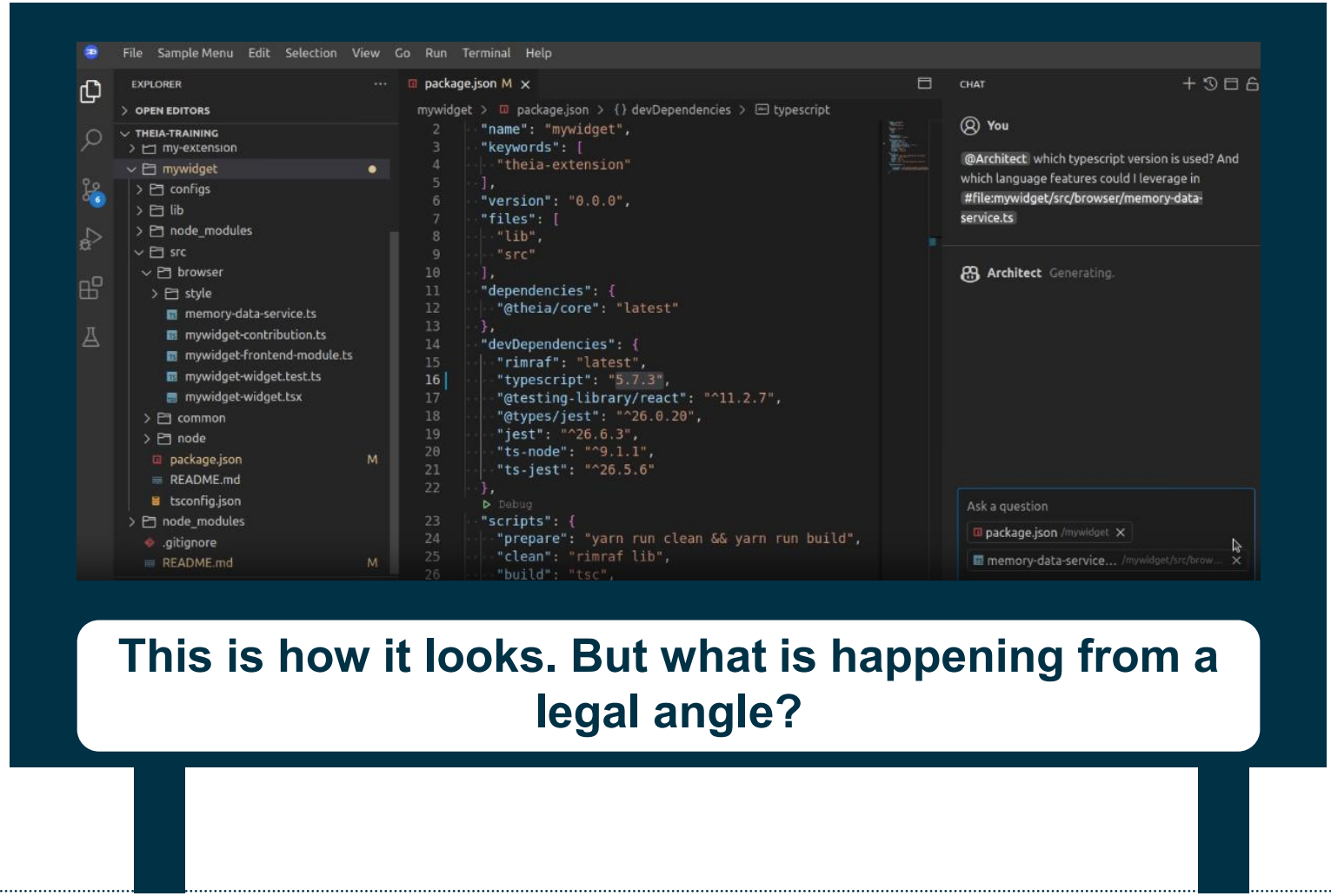
- generates docstrings, READMEs and comments in natural language
- explains code written by someone else or long-forgotten legacy code

Tools

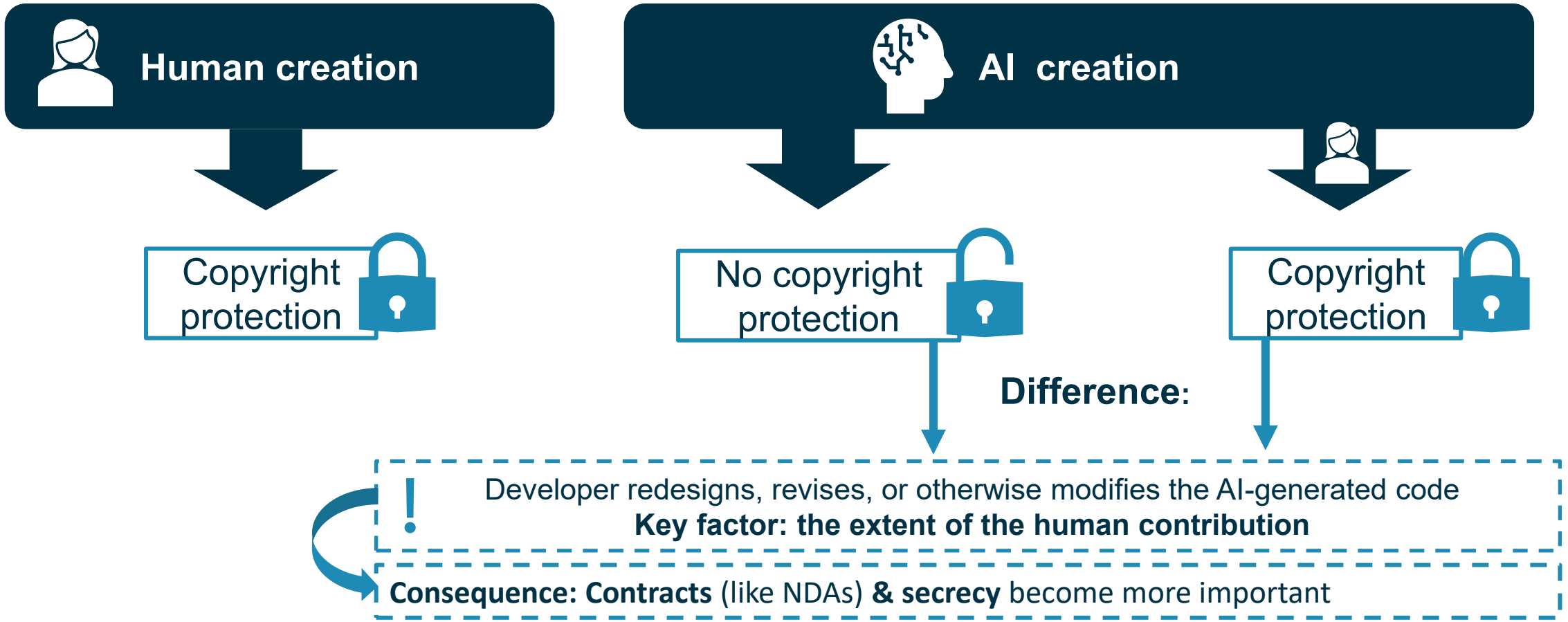
- GitHub Copilot
- Claude Code
- ChatGPT
- Theia IDE
- Cursor
- Windsurf

Different tools = different **training data**, **licenses**, and **IP/confidentiality** implications

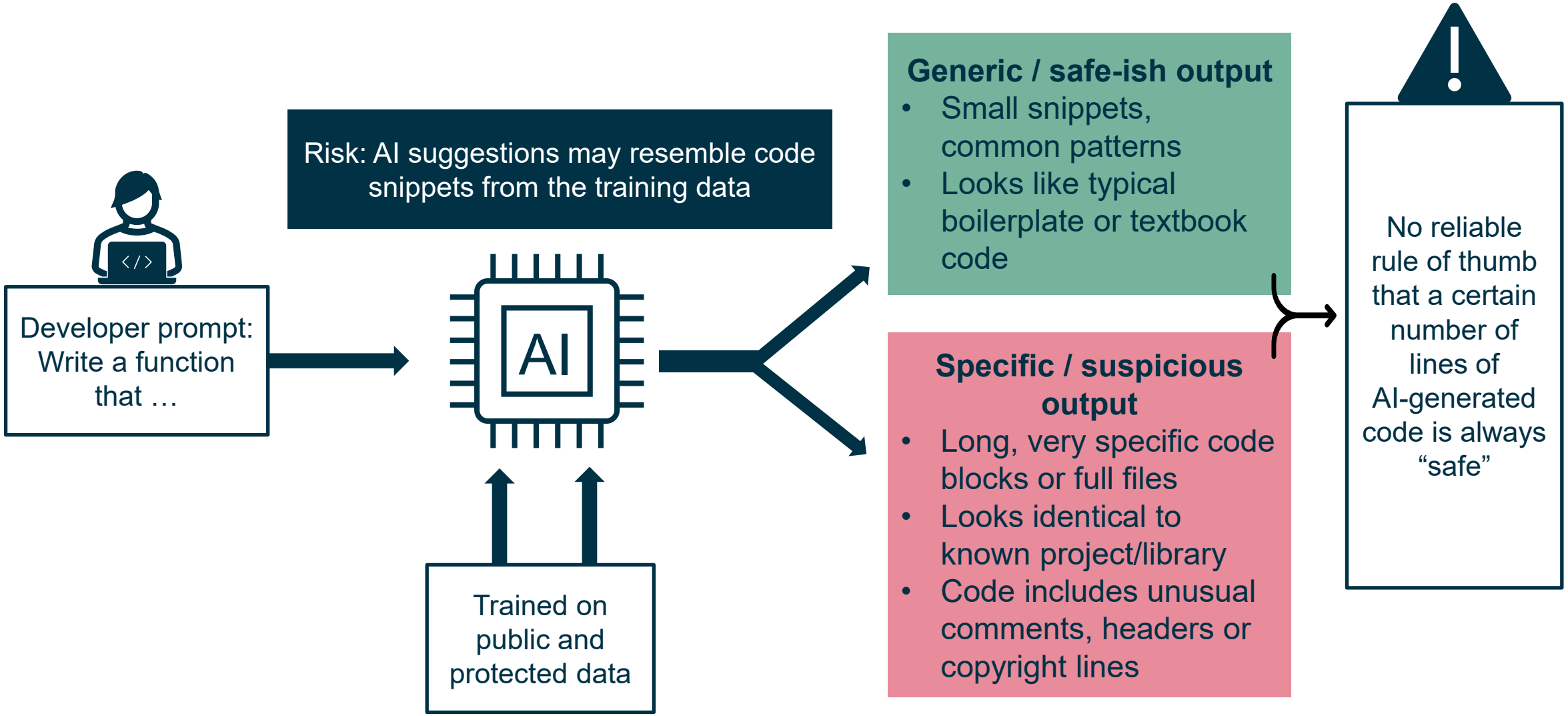
AI in everyday development – What are we already doing? (2)



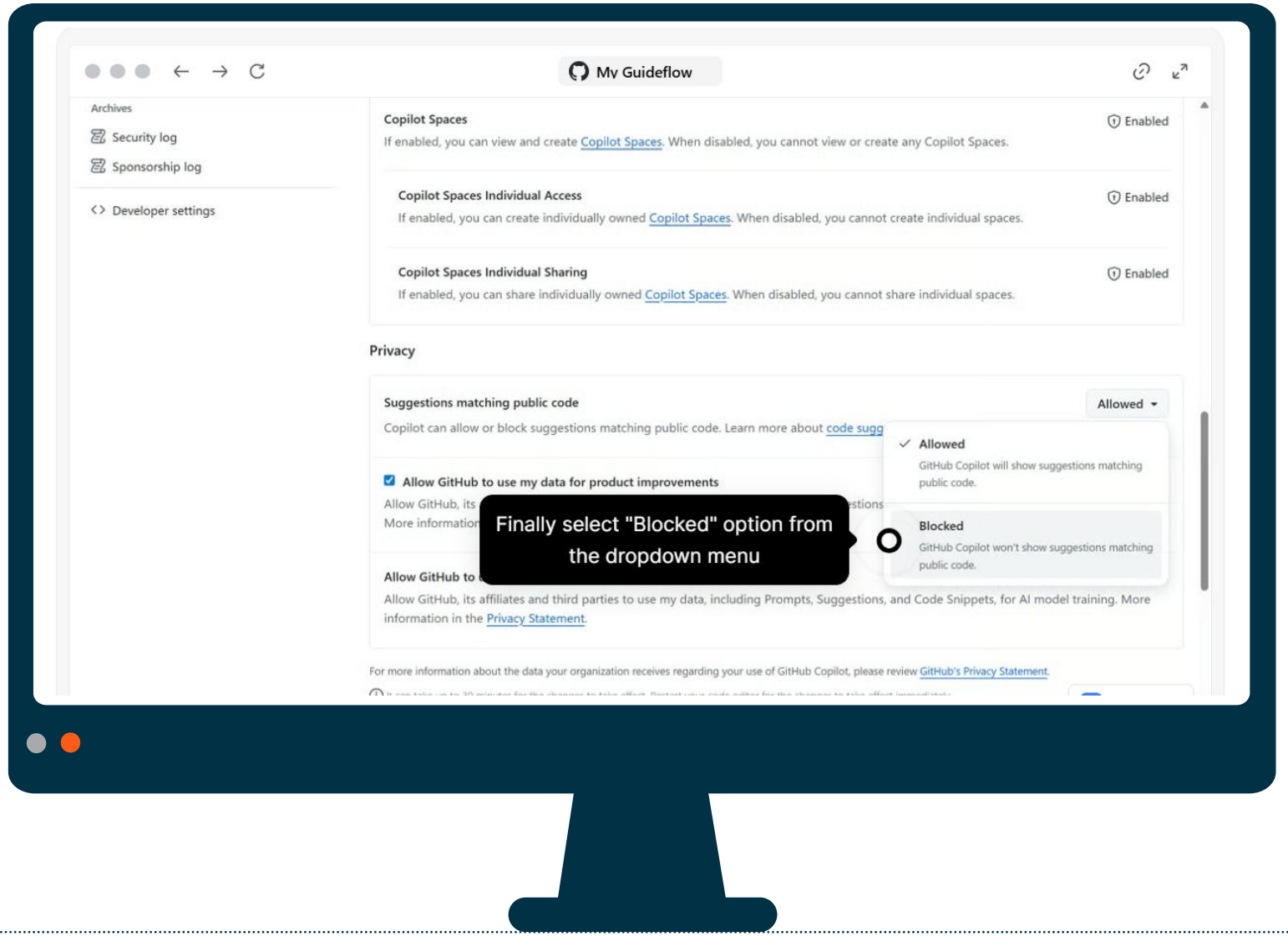
Copyright Protection for AI-Generated Code | Does AI Code Belong to Anyone?



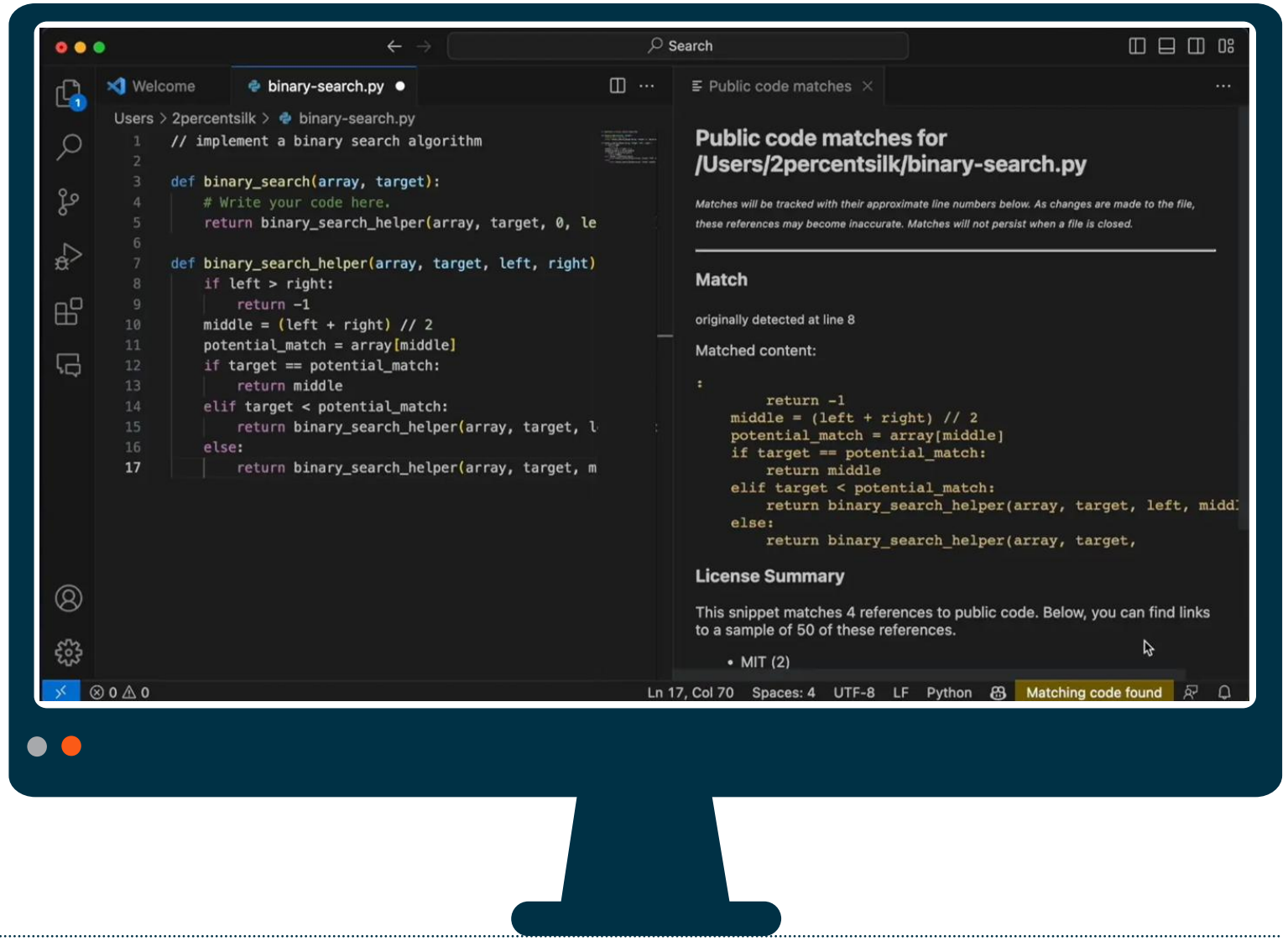
From Prompt to Code – When Output Becomes Risky



Filters for detecting public code duplicates



Code is scanned for patterns and watermarks and typically only reaches an accuracy of 20–25%.



How to minimize the chance of suspicious output

Use duplication (detection) filters built into services like GitHub Copilot

Check against publicly available repositories -> Snippet-Scanner

Minimising liability risk is crucial for damage claims and for providers to honour their copyright indemnity commitments

How to handle suspicious output

Treat as *suspect*: do a quick search / comparison

If in doubt, re-implement the logic in your own style

Avoid committing the snippet if you can't clarify its origin

Practical Do's & Don'ts

Dos

- **Choose AI platforms** whose capabilities, limitations and training sources **you understand**
- **Treat AI output like code from a junior dev** → review, refactor, test
- Be **extra careful with copyleft** (GPL/AGPL): check for "copy/paste-style" AI output
- Add a short **AI note to CONTRIBUTING/README** (e.g. "AI assistance allowed, but contributions must be legally clean")

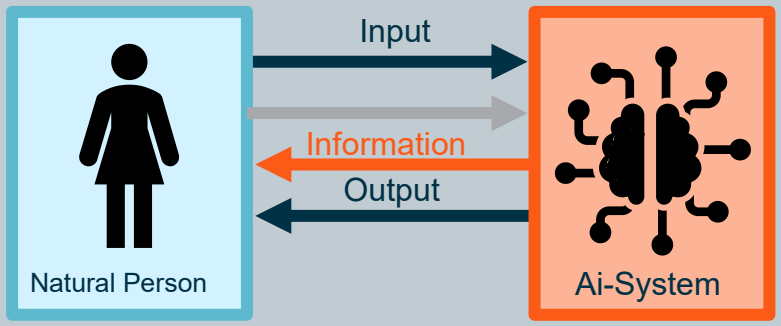
Don'ts

- Blindly commit long, very specific **AI-generated blocks** (especially in safety/security-critical components)
- Assume "AI wrote it, so it must be fine legally"
- **Forget to document** AI use (e.g. with a comment: "// Some portions generated by Co-Pilot")
- **Input confidential information**, personal data, health data or other sensitive information into AI
- **Use AI for prohibited purposes** (e.g. competitive uses) or in ways that conflict with your employer's policies

Regulatory Requirements | Transparency Obligation under Art. 50 (2) AI Act for Code??

Art. 50 (2) AI Act – Transparency Obligation

Source code can, according to the wording of the AI Regulation, indeed be subsumed under the concept of “synthetic text content”



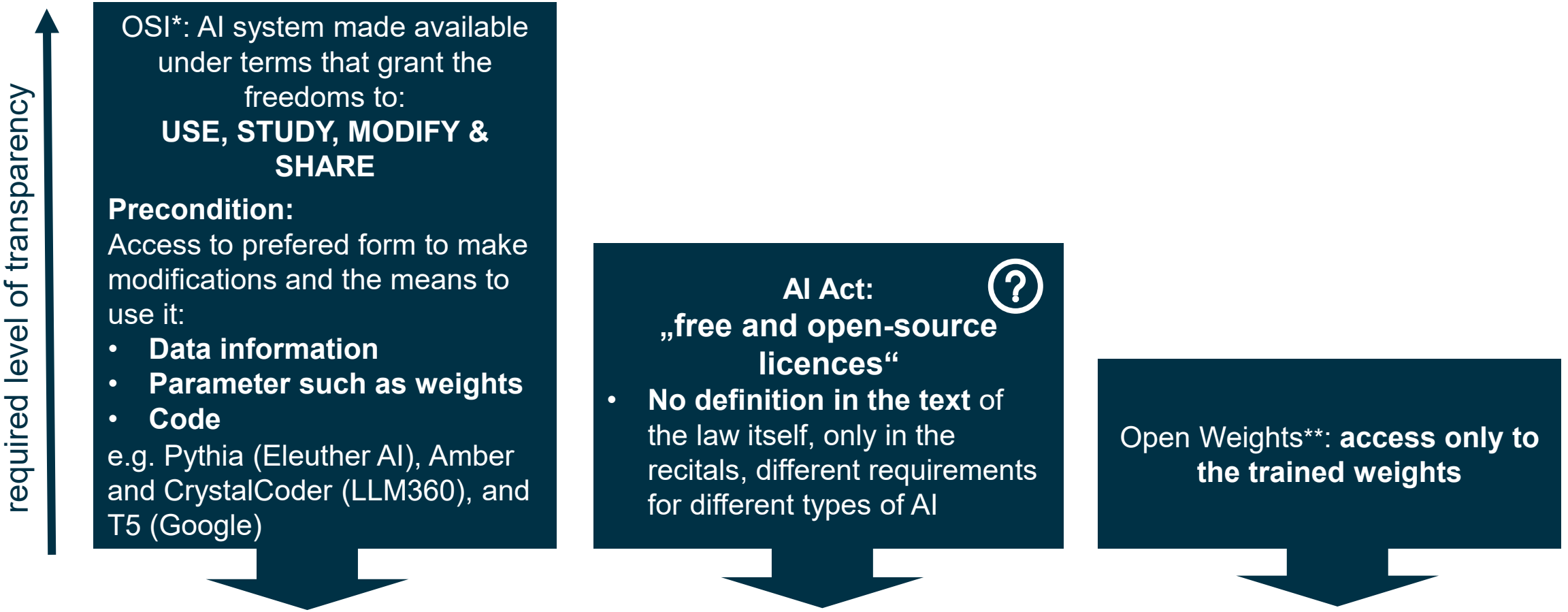
Consequence:

- According to the European Commission’s interpretation, source code as such is not intended to fall under the transparency obligation laid down in Article 50(2) AI Act.
- Reason: source code is not intended to be perceived, interpreted, verified or used as a basis for action by or for natural persons.

(Guidelines on the implementation of the transparency obligations for certain AI systems under Article 50 of Regulation (EU) 2024/1689 (the ‘AI Act’) Rn. 67 f.)

2 Use of Open Source AI

Open Source AI | What Is “Open Source AI”?



* available at: https://opensource.org/ai ** https://openweight.org

Open Source AI | Examples: OpenRail, Llama and OpenMDW Licences

DAMS

- OpenRAIL-D ▶ DATA
- OpenRAIL-A ▶ APPLICATION
- OpenRAIL-M ▶ MODEL
- OpenRAIL-S ▶ SOURCE CODE

Llama

- Meta AI's Llama licence was developed specifically for the use and distribution of its own language model Llama
- It aims to regulate the use of the model and ensure that it is used for responsible and ethical purposes.

https://logos-world.net/meta-logo/

OpenMDW

- Being developed **for the use and distribution** of AI models
- **Permissive Licence**
  → combines the provisions of the Apache-2.0 and MIT Open Source licences and adapts them to AI models

- ➢ **Use** "for any purpose" within the meaning of the OS AI definition does not (in part) exist in cases in which restrictions on use are affected.
- ➢ A **modification** "for any purpose" within the meaning of the OS AI definition cannot be made if the modifications are subject to restrictions on use.
- ➢ **Redistribution** "for any purpose" within the meaning of the OS AI definition is not permitted.

No „regular" open (source) licences according to the OSD and OS AI

Open Source AI | Examples: OpenRail

DAMS

- OpenRAIL-D ▶ DATA
- OpenRAIL-A ▶ APPLICATION
- OpenRAIL-M ▶ MODEL
- OpenRAIL-S ▶ SOURCE CODE

"[…]
You agree not to use the Model or Derivatives of the Model:

- In any way that violates any applicable national, federal, state, local or international law or regulation;
- For the purpose of exploiting, harming or attempting to exploit or harm minors in any way;
- To generate or disseminate verifiably false information and/or content with the purpose of harming others;
- To generate or disseminate personal identifiable information that can be used to harm an individual;

[…]"

✕ **No „regular" open (source) licence** according to the OSD and OS AI

Open Source AI | Examples: OpenMDW Licence (Draft)

OpenMDW Licence

- Being developed for the use and distribution of AI models
- Permissive licence
  → Combines the provisions of the Apache-2.0 and MIT Open Source Licences and adapts them to AI models

"[…]

As used in this agreement, "Model Materials" means the materials provided to you under this agreement, consisting of: (1) one or more **machine learning models (including architecture and parameters)**; and (2) **all related artefacts (including associated data, documentation and software)** that are provided to you hereunder. Subject to your compliance with this agreement, permission is hereby granted, **free of charge**, to deal in the Model Materials **without restriction**, including under all copyright, patent, database, and trade secret rights included or embodied therein.

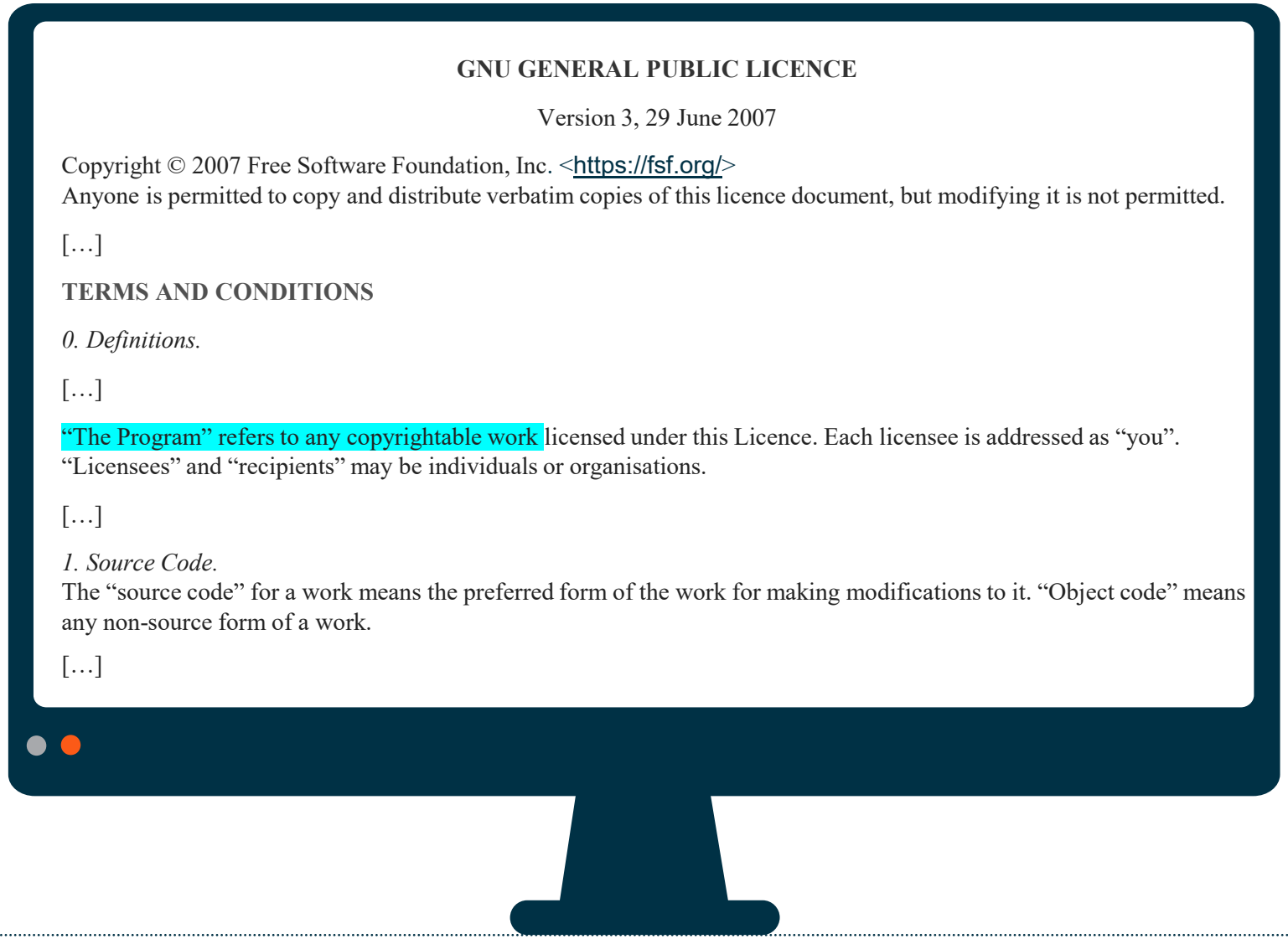
[…]

This agreement does **not impose any restrictions or obligations** with respect to **any use**, **modification**, **or sharing of any outputs generated by using the Model Materials**.

[…]"

Use / Modification / Distribution "**for any purpose**" → **Alignment with the OS AI**

Open Source AI | Example Model Licence (GPL 3.0)



GNU GENERAL PUBLIC LICENCE

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc. <https://fsf.org/>
Anyone is permitted to copy and distribute verbatim copies of this licence document, but modifying it is not permitted.

[…]

TERMS AND CONDITIONS

0. Definitions.

[…]

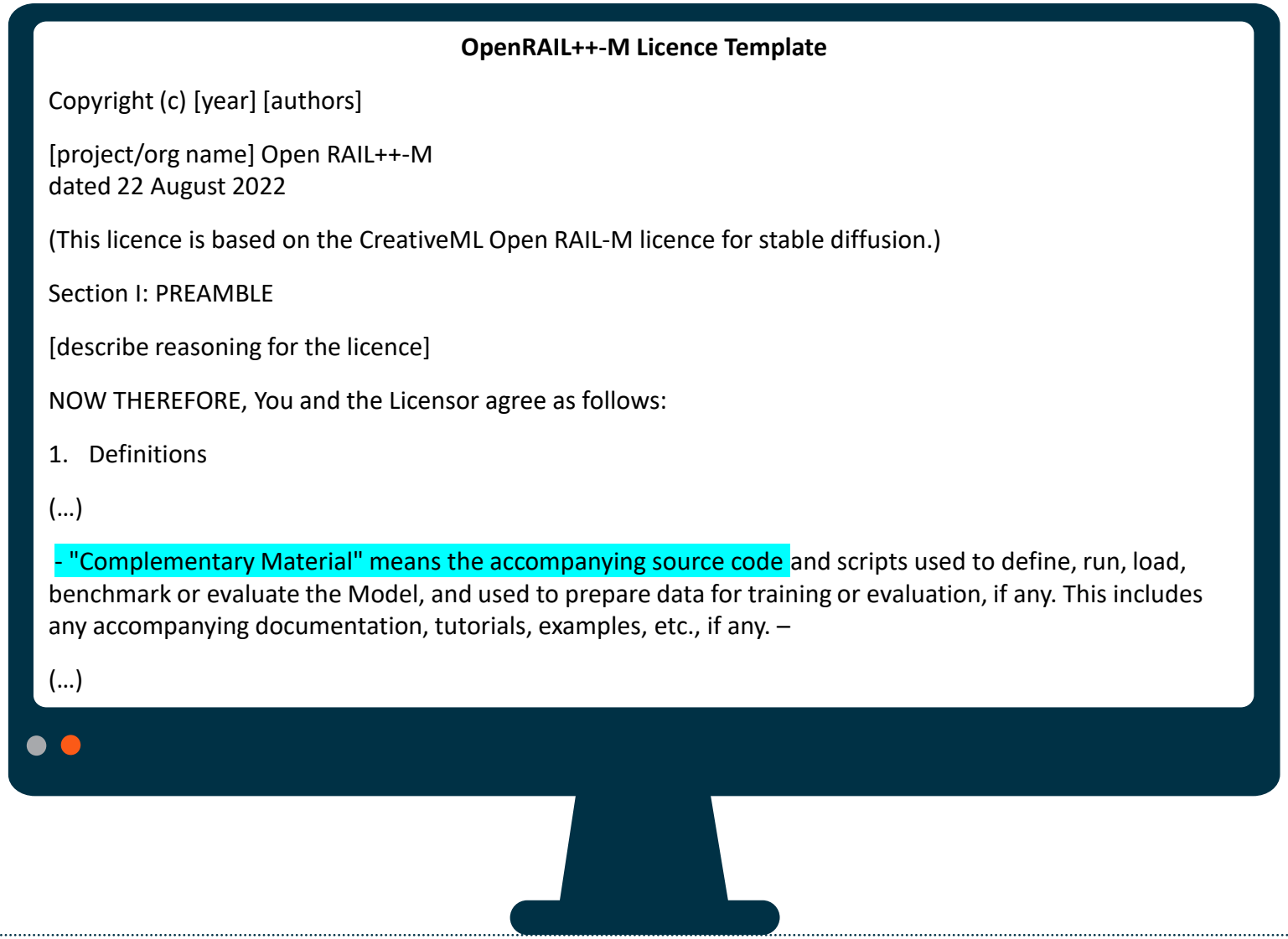
"The Program" refers to any copyrightable work licensed under this Licence. Each licensee is addressed as "you". "Licensees" and "recipients" may be individuals or organisations.

[…]

1. Source Code.
The "source code" for a work means the preferred form of the work for making modifications to it. "Object code" means any non-source form of a work.

[…]

Open Source AI | Example Code Licence (Open RAIL++-M)



OpenRAIL++-M Licence Template

Copyright (c) [year] [authors]

[project/org name] Open RAIL++-M
dated 22 August 2022

(This licence is based on the CreativeML Open RAIL-M licence for stable diffusion.)

Section I: PREAMBLE

[describe reasoning for the licence]

NOW THEREFORE, You and the Licensor agree as follows:

1. Definitions

(…)

- "Complementary Material" means the accompanying source code and scripts used to define, run, load, benchmark or evaluate the Model, and used to prepare data for training or evaluation, if any. This includes any accompanying documentation, tutorials, examples, etc., if any. –

(…)

Open-source AI | Conflict between model and code licences

Scope of AI protectability?

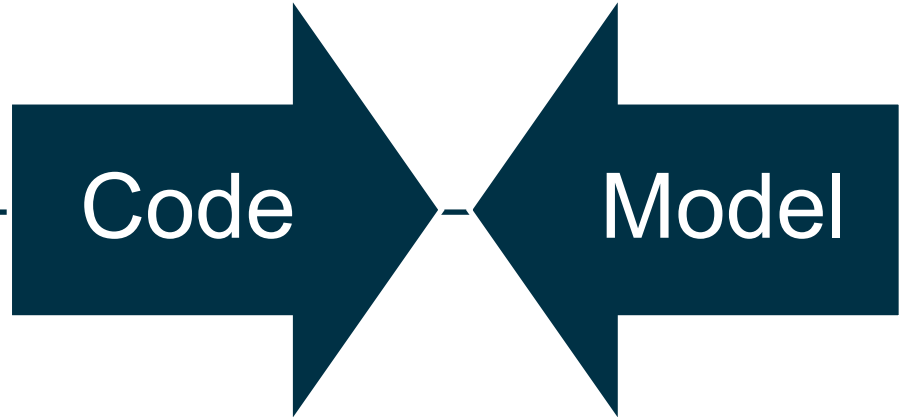
Overlap of scopes of application?

Q: Model Licence includes Source Code

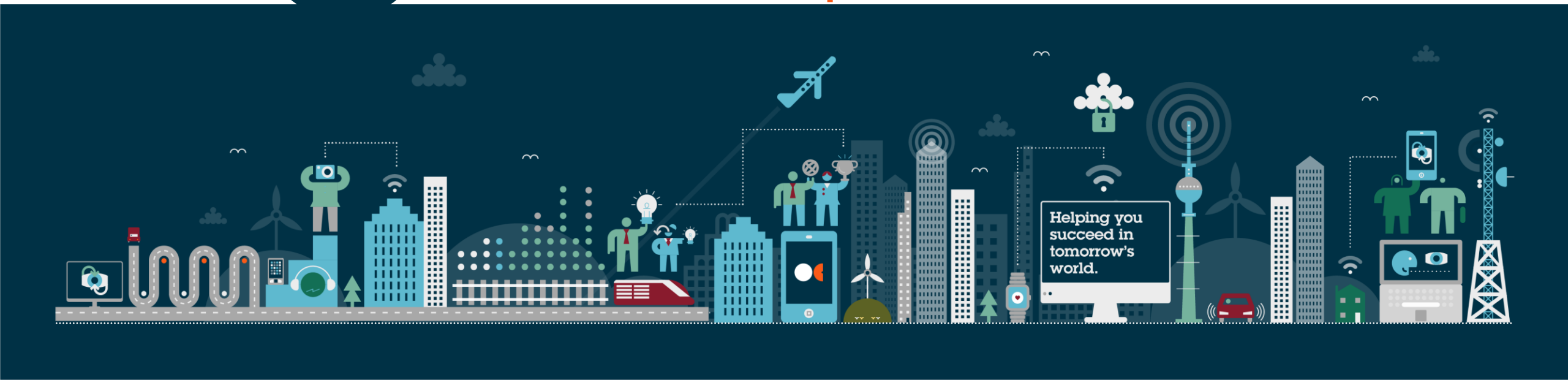
The Open RAIL++-M licence grants the right to use the model, derivatives of the model and supplementary material, **including the associated source code**.
However, it must be interpreted such that the **code licence itself can only apply to code that** is **not licensed under GPL-3.0**.

Mutual copyleft?

Sufficient granting of rights of use?



3 Common Legal Pitfalls with Open Source and Open Source AI & Solutions



Common Legal Pitfalls with Open Source and Open Source AI & Solutions

Open Washing:

- **What happens:** Assuming the label "Open Source" means completely open
- **Risk:** Hidden use restrictions (e.g. no certain industries) that break your planned business model
- **Practical fix:** Carefully **check the license**!

No Clear Origin of Models and Data

- **What happens:** Training/fine-tuning models without recording where the data/model came from or under which terms
- **Risk:** Hard to answer later: "What's in this model?"
- **Practical fix:** Maintain a very simple "**model bill of materials**"

Copyleft blind spots

- **What happens:** Using GPL/AGPL/ EPL
- **Risk:** Unintended obligations to publish source code or even parts of your service stack, especially in SaaS scenarios
- **Practical fix:** Flag early, avoid them in core services where possible, **or get a targeted legal check**

Ignoring the AI Act/ GDPR etc.

- **What happens:** Assuming these Acts don't apply because the model is "open source"
- **Risk**: Overlooking duties e.g. transparency
- **Practical Fix**: Do a simple **GDPR/AI Act check early on**

Thank you!

Lina Böcker
Partner
Germany

+49 30 72621 8095
lina.boecker@osborneclarke.com