

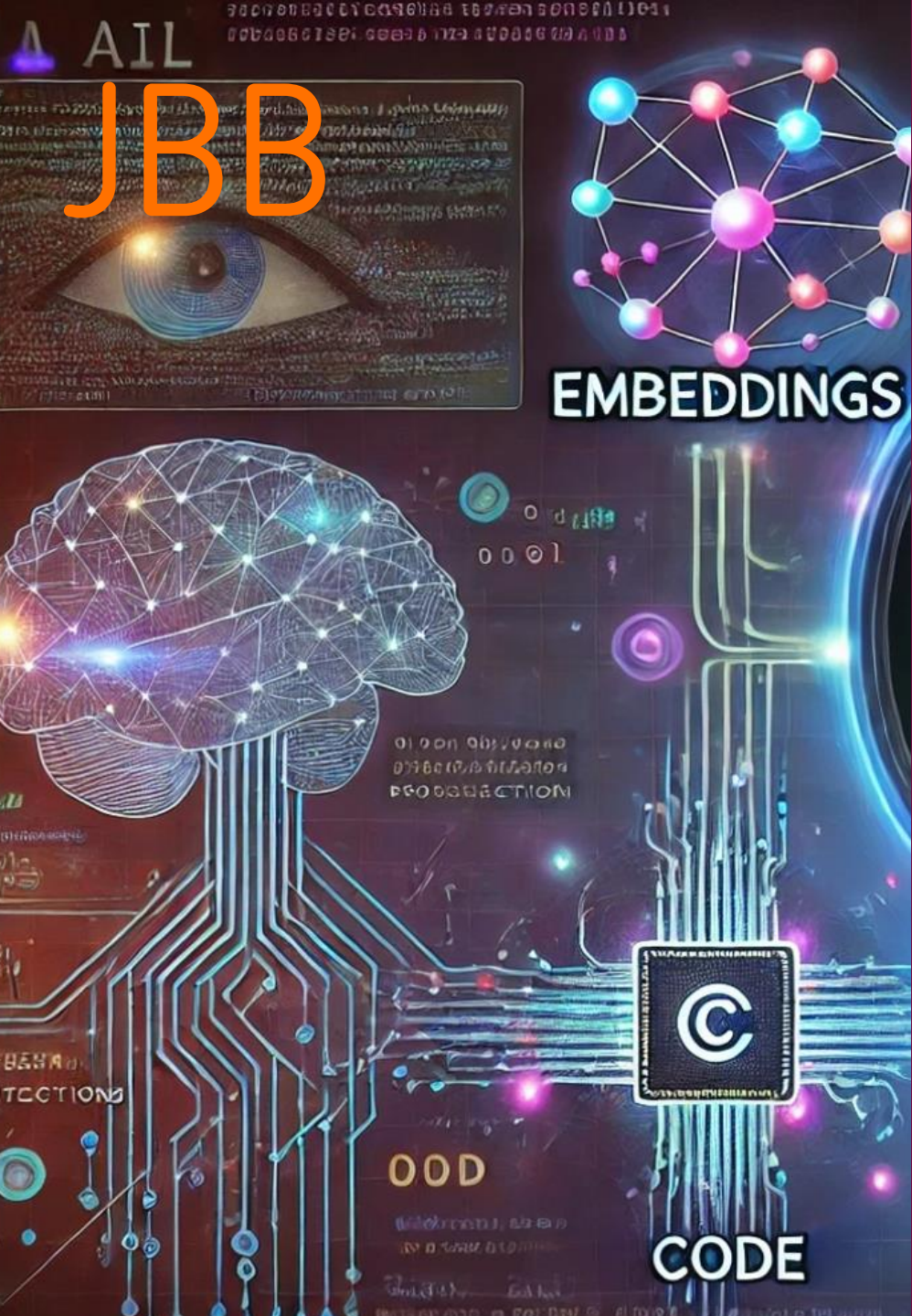
Copyright Issues With Training & Using AI Models

Dr. Lisa Käde

JBB Rechtsanwält:innen

26.03.2025

COOL COMPACT
OSADL
ONLINE
LECTURES

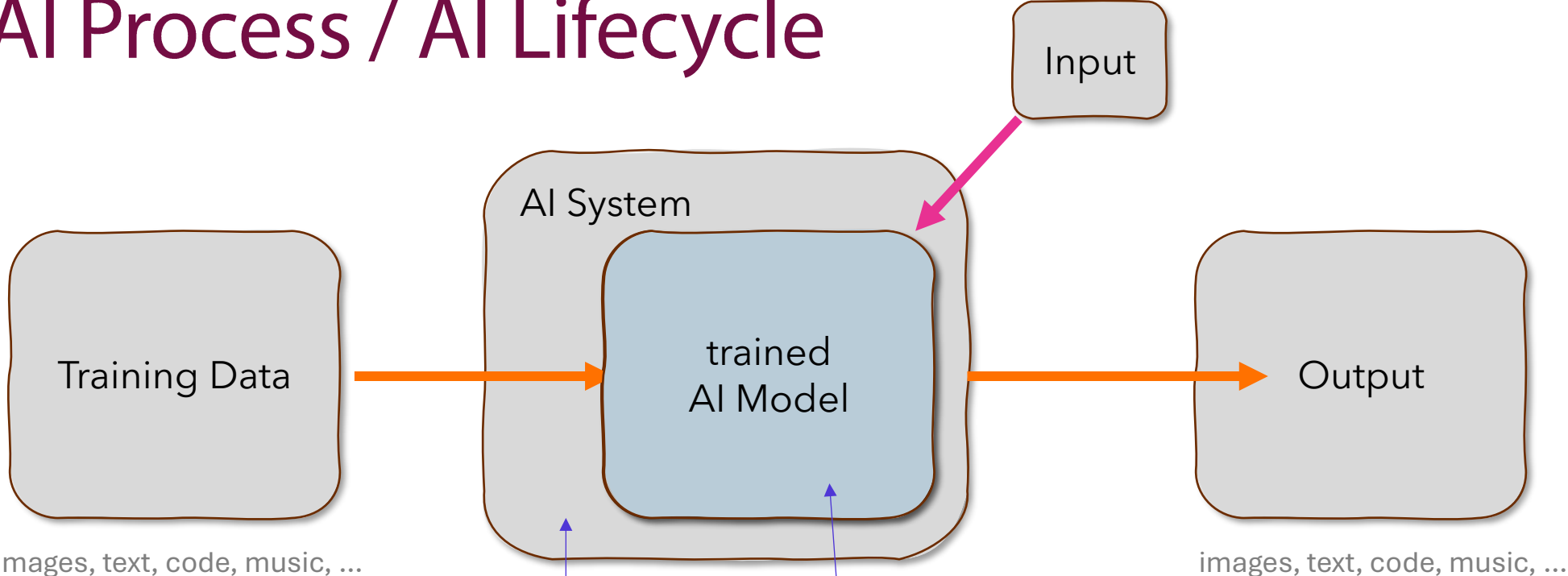


AI / Machine Learning Models

The Essence of AI Models – Assumptions

- No traditional computer programs, but **controlled** by them and often also **part** of them
- (Embeddings +) Hyperparameters + Parameters + Code = Model
- Model + Training Data = Trained Model
- Trained Model + Input = Output

The AI Process / AI Lifecycle

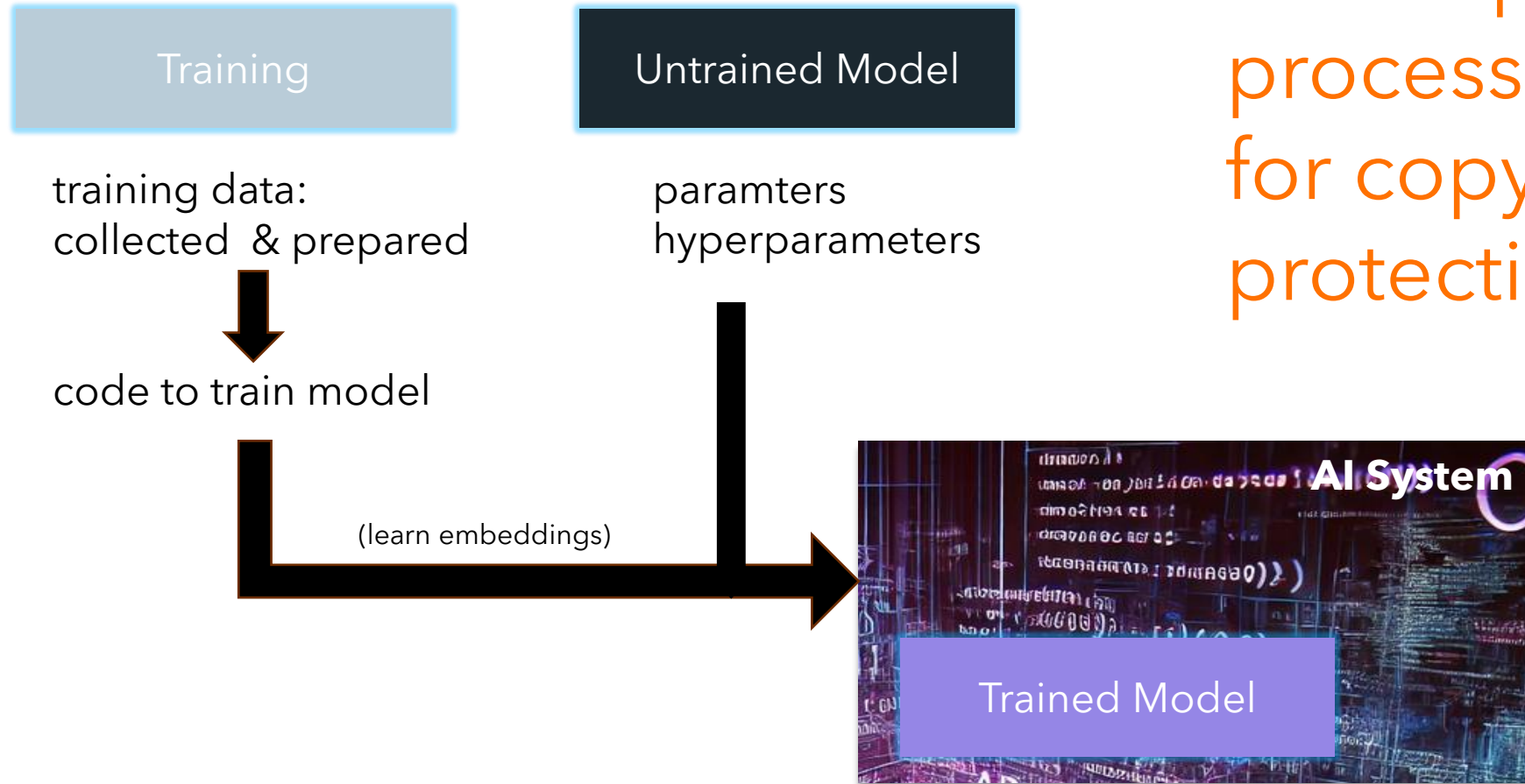


copyright for individual items & for database ✓

copyright for code ✓

copyright for ... ?

Overly simplified depiction of AI training process

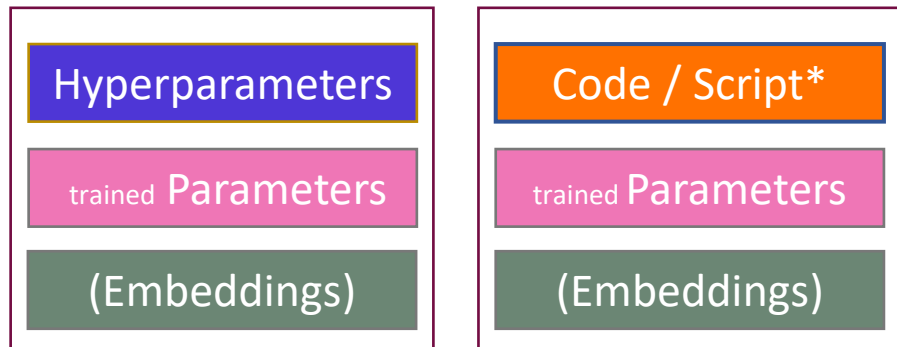


Which parts of this process are relevant for copyright protection?

Model Protection What is the essence that should be (copyright) protected?

- It is worth to protect whatever is the result of the efforts of the developer (investment of time and resources)
- It is worth to protect whatever can easily replicate the solution to the respective problem
- It is worth to protect whatever constitutes the competitive advantage of the company

Minimum requirements for replication:



* including hyperparameters

Assumptions (again):

- Parameters without hyperparameters are useless
- Hyperparameters without parameters OR training data are useless
- Code without parameters, hyperparameters & training data is useless

What protection does copyright offer? (in DE)

Computer program protection
(§§ 69a ff. UrhG)

computer program
controlling commands

own intellectual creation

Database work
(§ 4 UrhG)

collection (work)
independent elements
systematic / methodic ordering
individually accessible

personal intellectual creation
ordering
selection

Database sui generis
(§ 87a UrhG)

collection
independent elements
systematic / methodic ordering
Individually accessible

relevant investment
in data acquisition / presentation / validation

Other work
(§ 2 UrhG)

work character

personal intellectual creation

Trade Secrets
(not copyright)

Protection – Status Quo (?) [excl. Embeddings]

	Hyperparameters	Parameters	Code / Script	All-In-One
--	-----------------	------------	---------------	------------

database work

✓ ** / ✗	✗	✗	✗	✓ ***
----------	---	---	---	-------

database sui generis

✓	✓ ** / ✗	✗	✗	✓ ***
---	----------	---	---	-------

computer program protection

✗	✗	✓ ** / ✗	✗	✓ ** / ✗
---	---	----------	---	----------

** depending on

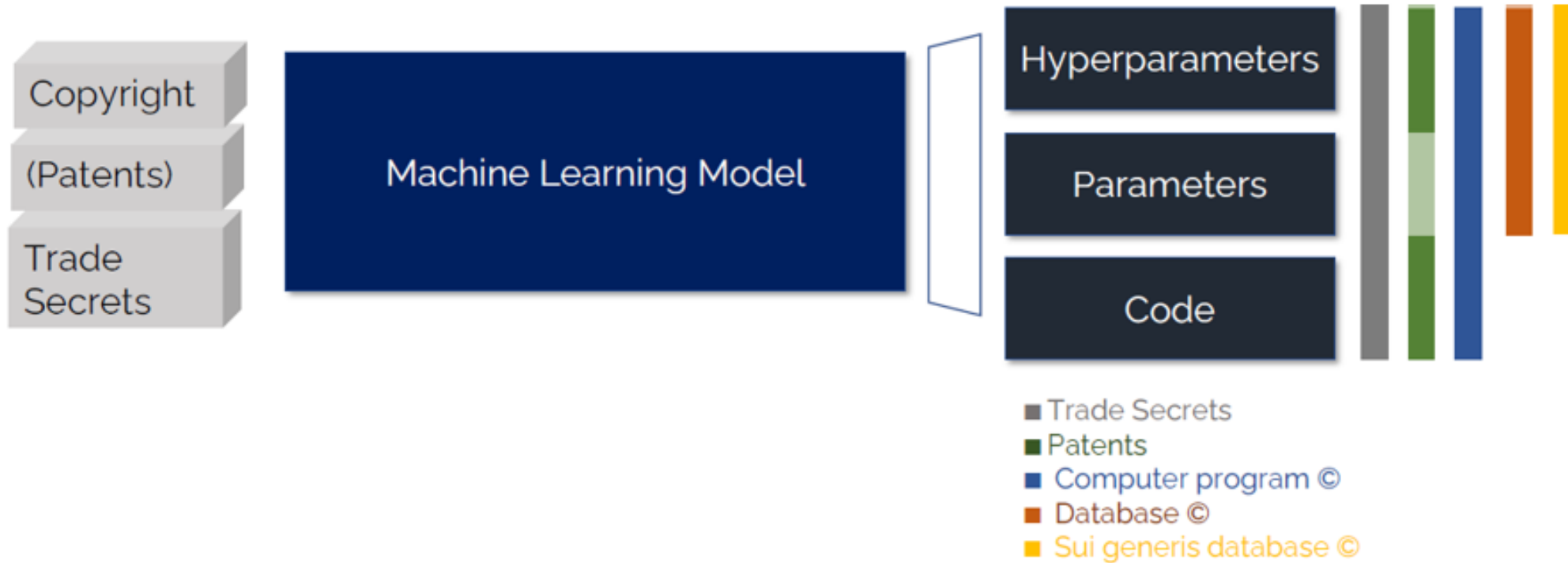
individuality of hyperparameters

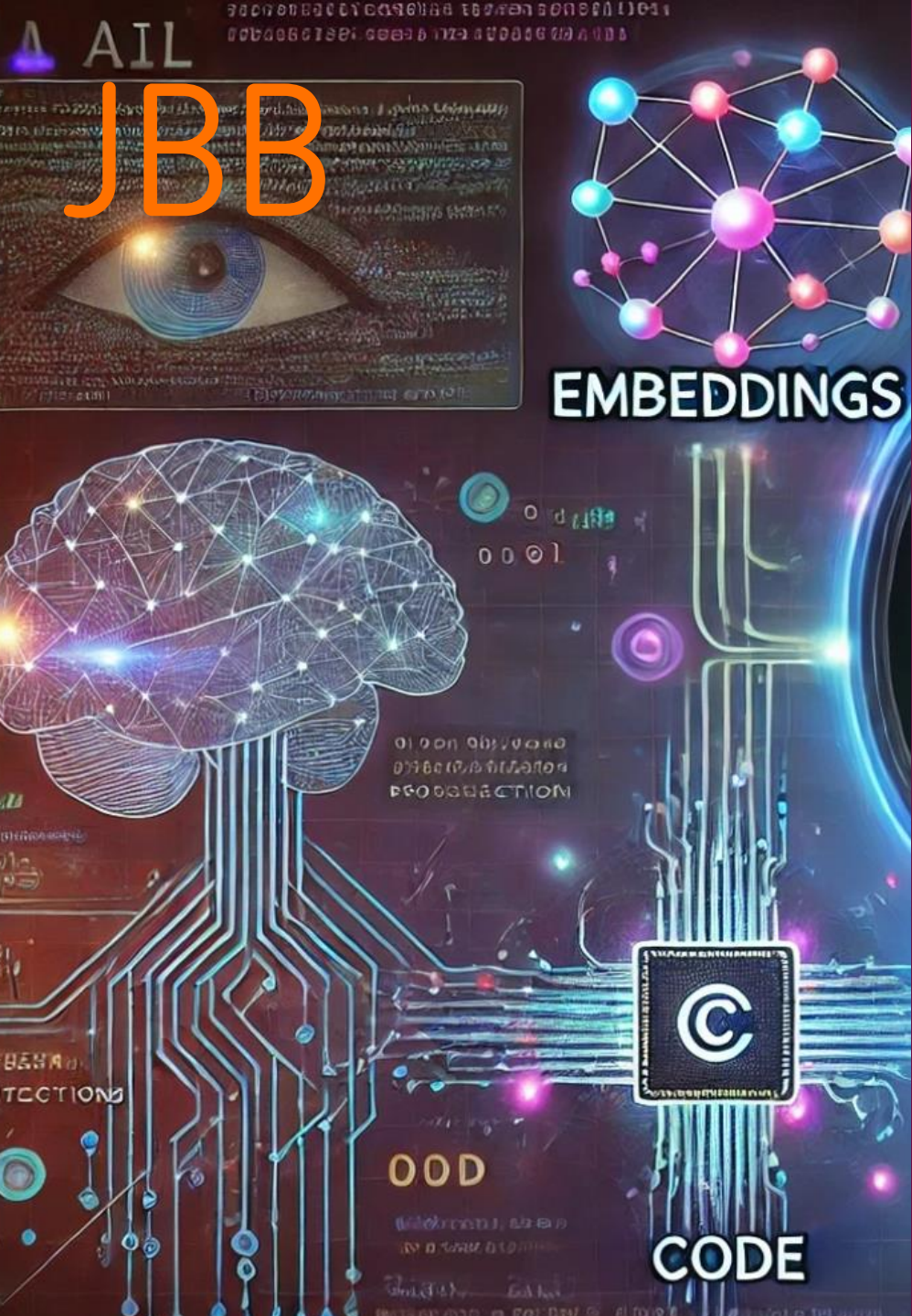
type of AI system

individuality of source code - „Building Blocks“

*** protection not extending to source code

Protection – Status Quo (?) [excl. Embeddings]

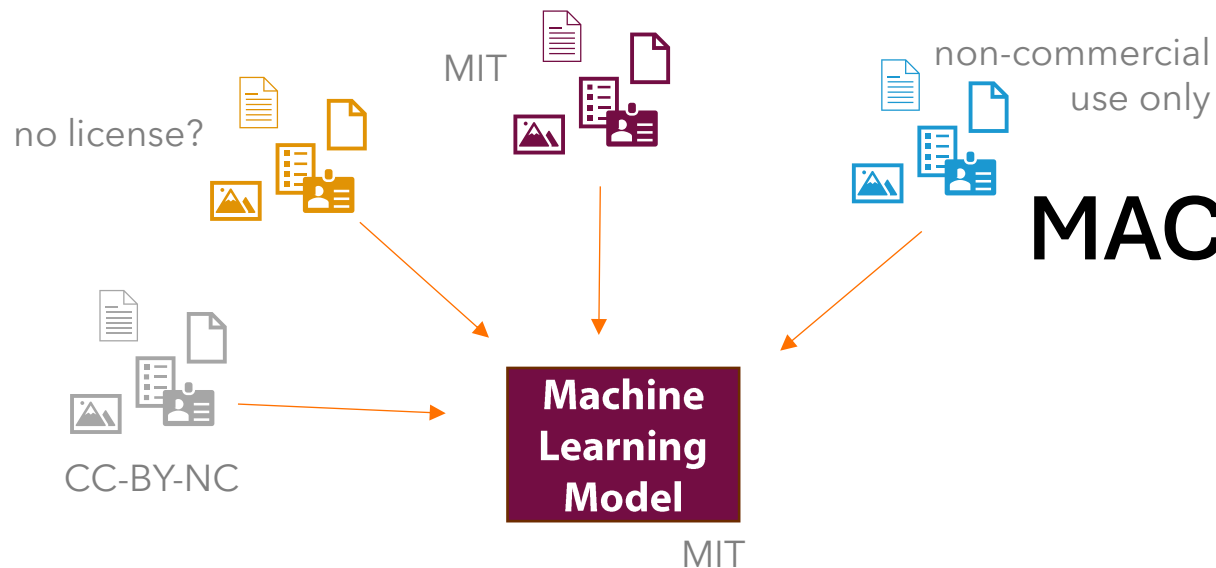




Licensing of AI Models

TRAINING DATA AND MODEL MAY BE SUBJECT TO DIFFERENT COPYRIGHT ASPECTS AND DIFFERENT LICENSES.

DIFFERENT TRAINING DATA SETS MAY BE USED TO TRAIN ONE MACHINE LEARNING MODEL

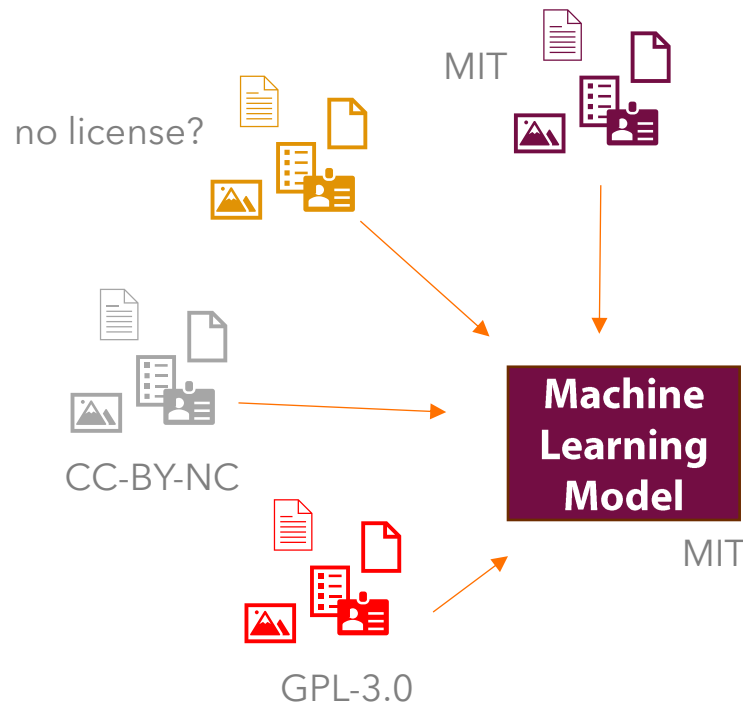


DIFFERENT

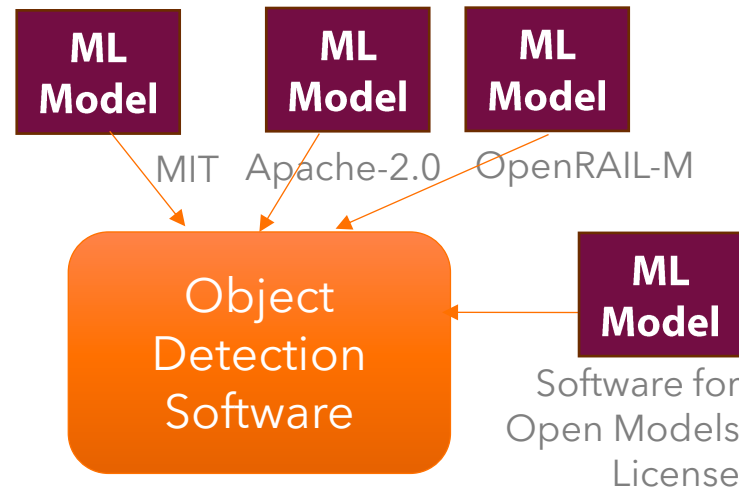
**TRAINING DATA SETS
MAY BE USED TO TRAIN**

ONE

MACHINE LEARNING MODEL



**MULTIPLE DIFFERENT
MODELS MAY BE
USED AS PART OF THE
NEW SOFTWARE**



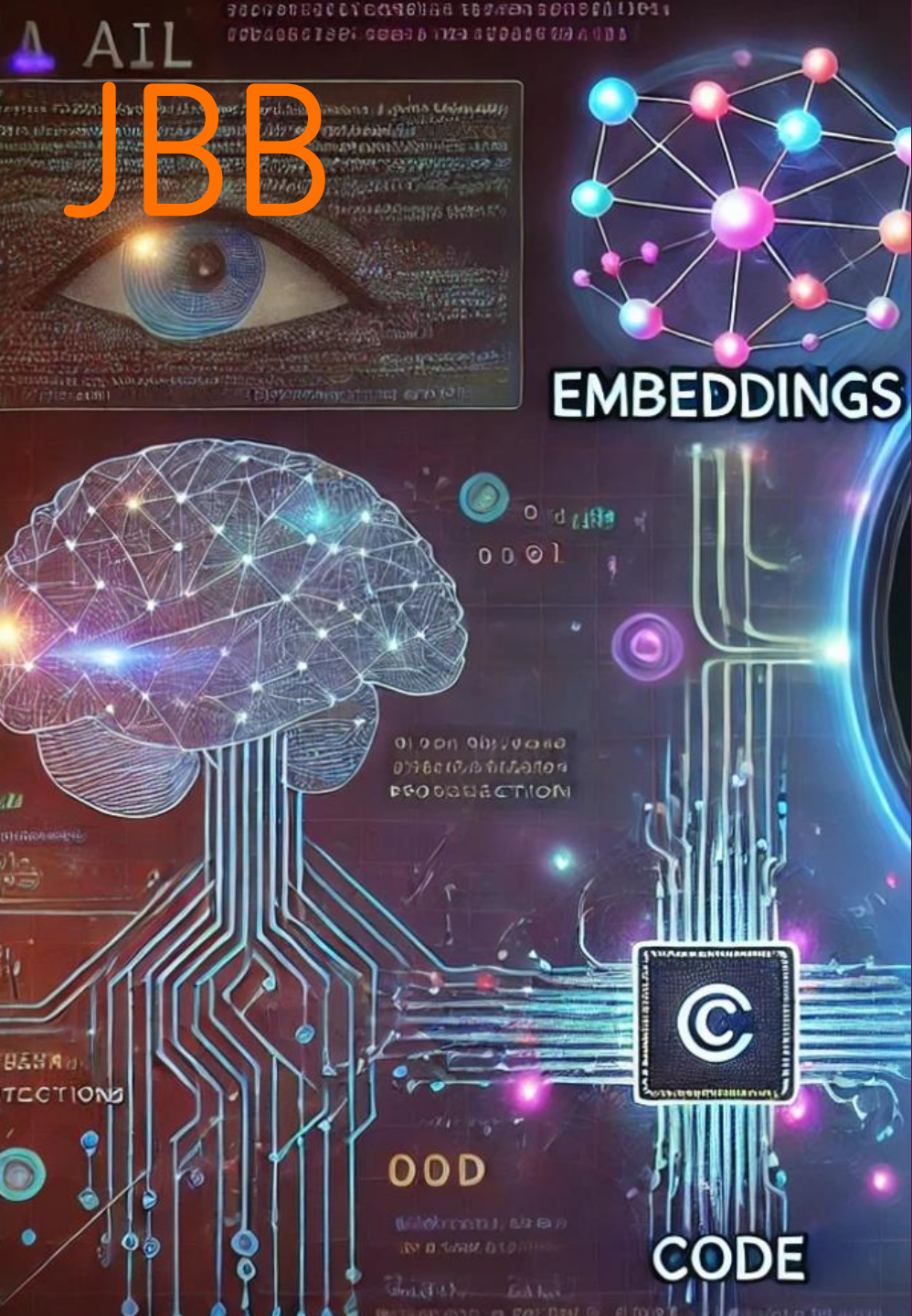
Potential Influence on Model Licensing

- + **Training data?** → does the model „contain“ training data – derivative work?
- + **Training code?** → is the model a derivative work of training code?
- + **Embeddings?** → how entangled are Model and Embeddings / Embedding Model?
- + **Use of pre-trained / fine-tuning of models?**
→ can the license be changed for a re-trained model?

Influence of Model Licensing on Use in Products

- + Licensing obligations apply like for other software
- + Copyleft effect? → potentially depending on use (API / integration)
- + Licensing might also (indirectly) affect AI Act obligations
- + AI licenses may restrict certain use cases („ethical licenses“)

- + **watch out:** licenses for code and model may differ



Open Source AI Definition

What is „Open Source AI“?

- use
- study
- modify
- share

- AI system
- model
- weights / parameters
- other structural elements

OS AI Definition (OSI),
opensource.org



- Information on **training data**
- **Code** (for training, data preparation, validation, test etc.)
- **parameters & configuration**



Open Source AI?

DeepSeek-R1 is now live and **open source**, rivaling OpenAI's Model o1. Available on web, app, and API. Click for details.

deepseek

Into the unknown

<https://huggingface.co/deepseek-ai/DeepSeek-R1/tree/main>

MIT License:

- code ✓
- weights (tensors) ✓
- configuration ✓

- most „Open Source“ AI models only open source their code
- model is released under license agreement restricting use (!= Open Source)
- or information about training data is missing

Llama

The open-source AI models you can fine-tune and deploy anywhere. Choose from our collection of Llama 3.1, Llama 3.2, Llama 3.3.

Download models

Try Llama on Meta

Text **New**

Llama 3.3: 70B

- State-of-the-art multilingual open source large language model
- Experience 405B performance and quality at a fraction of the cost

*Licensed under Llama 3.3 Community License Agreement

Recap

- + ML models may be copyrightable subject matter
- + Licensing obligations need to be observed
- + Training data licenses most likely do not influence the model license

- + **Which training data may be used for training at all?**
That's up next!



JBB

JBB Rechtsanwälte
Jaschinski Biere Brexl Partnerschaft mbB

Dr. Lisa Käde, B.Sc.
Rechtsanwältin

Christinenstraße 18/19
10119 Berlin

Tel. +49 30 443 765 0

Mail kaede@jbb.de

Web www.jbb.de