Building the world's first Dictionary of Applied Machine Learning

Alexander Jung
Aalto University
alexjung.at

FLAIG Summit · 3 November 2026

BUSINESS FINLAND

Forward-Looking AI Governance in Banking and Insurance (FLAIG)
Business Finland co-research project 7360/31/2023 · 1 October 2024–30 September 2026
University of Turku (coordinator), Aalto University, OP Financial Group, Veritas, Tietoevry

Supervisors are already acting

GDPR penalties against financial institutions that are final and enforceable

€5 million

BBVA · AEPD, December 2020 · upheld by the Supreme Court, November 2024

Unclear information about processing, and an invalid mechanism for obtaining customer consent.

SEK 7.5 million

Klarna Bank · IMY, March 2022 · restored on appeal, March 2024

Incomplete and unclear information about the processing and about the rights of data subjects. About €660,000.

€492,000

Credit provider, not named · Hamburg DPA, announced 30 September 2025 · notice accepted

Credit-card applications refused by fully automated decision, with no meaningful account of the logic.

The last of these is a failure of explanation, not of security.

The ceilings

GDPR, Article 83

€10m or 2%	controller and processor duties, security, breach notification
€20m or 4%	the basic principles, lawful basis, data subject rights, transfers

AI Act, Article 99

€35m or 7%	the prohibited practices of Article 5
€15m or 3%	most other duties, high-risk systems included
€7.5m or 1%	misleading information to authorities

The ceiling is measured on the whole group's worldwide annual turnover.

The courts now demand an explanation

Court of Justice of the European Union · automated credit assessment

C-203/22 — **Dun & Bradstreet Austria** · 27 February 2025

"Meaningful information about the logic involved" means explaining the procedure and principles actually applied, concisely and intelligibly. Disclosing the algorithm does not satisfy the duty.

C-634/21 — **SCHUFA Holding (Scoring)** · 7 December 2023

An automated probability value on the ability to service a loan is itself a decision under Article 22, where the bank's conduct depends decisively on that value.

The Court requires an intelligible explanation. It does not say how to produce one.

The law begins with understanding

Regulation (EU) 2024/1689 (AI Act), Article 4

> "Providers and deployers of AI systems shall take measures to support the development of **AI literacy** of their staff and other persons dealing with the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in ..."

Literacy presupposes a shared vocabulary.

One AI system, three professional languages

Developer

"Gradient boosting on tabular features, tuned on a validation set."

Compliance

"Is it transparent, fair and privacy-preserving?"

Regulator

"Is it an AI system? Who is the provider?"

Each group is fluent, but their claims are difficult to compare or verify.

A bank has to satisfy all three at once.

Governance needs a working language

- Standards and the EU AI Act begin with terms and definitions.
- Legal definitions establish duties, but do not say how to measure a trained model.
- A working language connects law, compliance and machine-learning practice.

One definition per term — written down, cross-referenced and checkable.

Three rules for every entry

1. One picture

Every term is anchored on the same figure: data points, a model drawn as two candidate hypothesis maps, and the loss between them.

2. Cross-references

Every technical dependency points to its own entry.

3. A runnable check

A Python demo recomputes the claims and reports pass or fail.

A definition you can run is a definition you can audit.

Two artefacts, one source

dictionaryofml.org

the open site

- 114 entries, free to read and reuse
- CC BY 4.0, runnable Python demos
- machine-readable, queryable by a coding agent
- revised continuously

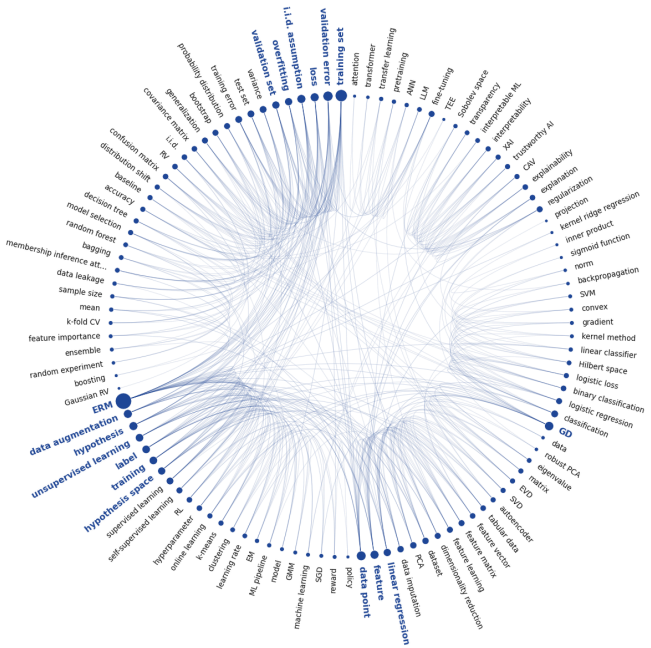
Dictionary of Applied Machine Learning

the Springer reference work

- peer-reviewed and editorially controlled
- each entry separately versioned and citable
- a stable record for reports and audits
- publishing entry by entry since 1 November 2026

The site moves fast so that practice stays current. The reference work holds still so that an audit can cite it.

The open site



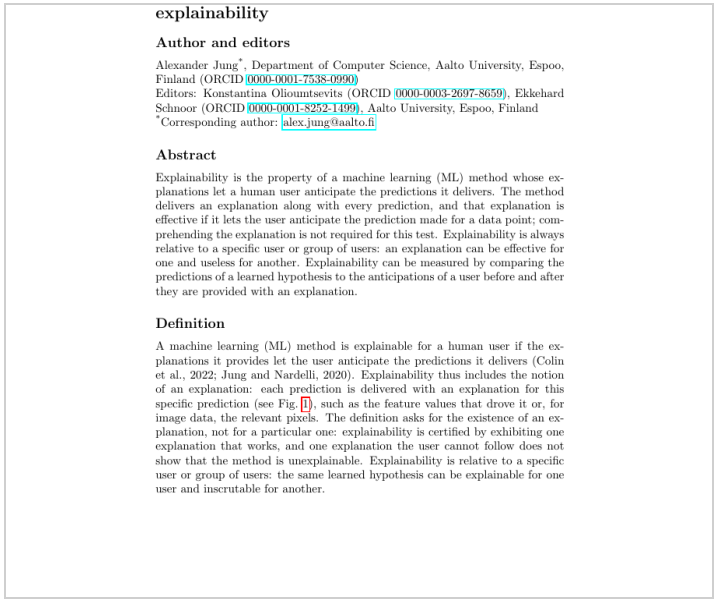
dictionaryofml.org

114 entries, 437 cross-references, CC BY 4.0, archived at DOI 10.5281/zenodo.21569296.

One server, three kinds of client

a reader	a typeset HTML page, one per term
a machine	terms.json and a Colab notebook that recomputes the claims
a coding agent	the MCP endpoint at dictionaryofml.org/mcp

Three perspectives on each term



explainability

Author and editors

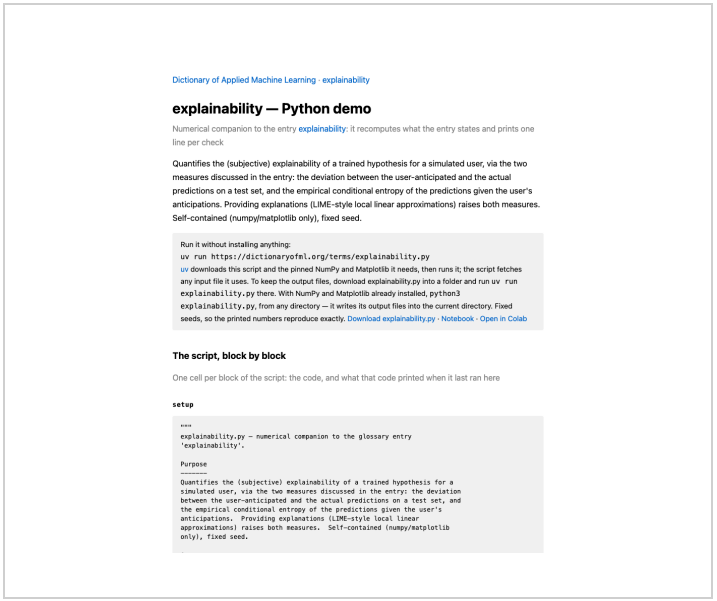
Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland (ORCID 0000-0001-7538-0990)
Editors: Konstantina Olioumtsevits (ORCID 0000-0003-2697-8659), Ekkehard Schnoor (ORCID 0000-0001-8252-1499), Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Abstract

Explainability is the property of a machine learning (ML) method whose explanations let a human user anticipate the predictions it delivers. The method delivers an explanation along with every prediction, and that explanation is effective if it lets the user anticipate the prediction made for a data point; comprehending the explanation is not required for this test. Explainability is always relative to a specific user or group of users: an explanation can be effective for one and useless for another. Explainability can be measured by comparing the predictions of a learned hypothesis to the anticipations of a user before and after they are provided with an explanation.

Definition

A machine learning (ML) method is explainable for a human user if the explanations it provides let the user anticipate the predictions it delivers (Colin et al., 2022; Jung and Nardelli, 2020). Explainability thus includes the notion of an explanation: each prediction is delivered with an explanation for this specific prediction (see Fig. 1), such as the feature values that drove it or, for image data, the relevant pixels. The definition asks for the existence of an explanation, not for a particular one: explainability is certified by exhibiting one explanation that works, and one explanation the user cannot follow does not show that the method is unexplainable. Explainability is relative to a specific user or group of users: the same learned hypothesis can be explainable for one user and inscrutable for another.



Dictionary of Applied Machine Learning · explainability

explainability — Python demo

Numerical companion to the entry explainability: it recomputes what the entry states and prints one line per check

Quantifies the (subjective) explainability of a trained hypothesis for a simulated user, via the two measures discussed in the entry: the deviation between the user-anticipated and the actual predictions on a test set, and the empirical conditional entropy of the predictions given the user's anticipations. Providing explanations (LIME-style local linear approximations) raises both measures. Self-contained (numpy/matplotlib only), fixed seed.

Run it without installing anything:
uv run https://dictionaryofml.org/terms/explainability.py
uv downloads this script and the pinned NumPy and Matplotlib it needs, then runs it; the script fetches any input file it uses. To keep the output files, download explainability.py into a folder and run uv run explainability.py there. With NumPy and Matplotlib already installed, python3 explainability.py, from any directory — it writes its output files into the current directory. Fixed seeds, so the printed numbers reproduce exactly. Download explainability.py · Notebook · Open in Colab

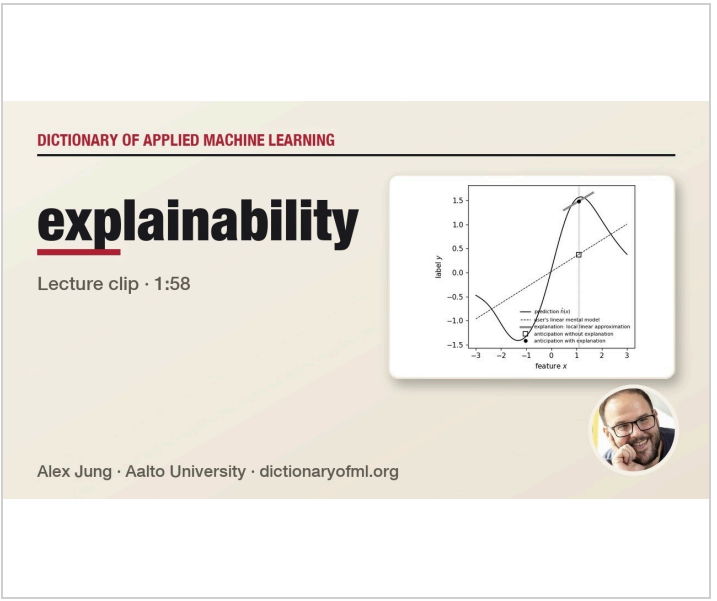
The script, block by block

One cell per block of the script: the code, and what that code printed when it last ran here

setup

```
"""
explainability.py — numerical companion to the glossary entry
'explainability'.

Purpose
-------
Quantifies the (subjective) explainability of a trained hypothesis for a
simulated user, via the two measures discussed in the entry: the deviation
between the user-anticipated and the actual predictions on a test set, and
the empirical conditional entropy of the predictions given the user's
anticipations.  Providing explanations (LIME-style local linear
approximations) raises both measures.  Self-contained (numpy/matplotlib
only), fixed seed.
```



PDF / HTML

the typeset entry

Python demo

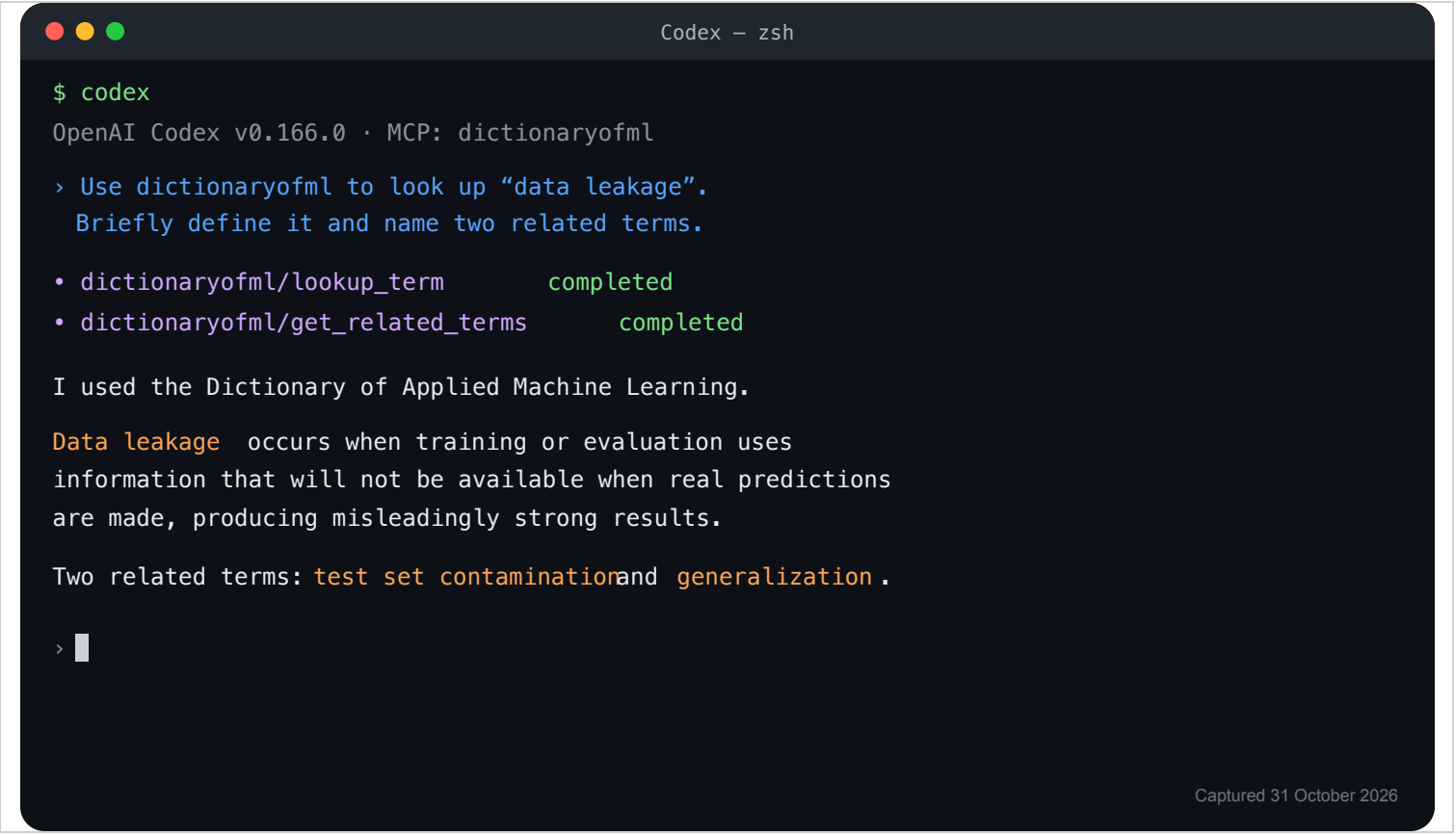
recomputes the claims

YouTube Short

two minutes

One source feeds all three, and the Springer submission.

Serving a coding agent



Actual Codex CLI lookup, captured 31 October 2026.

1. Ask

Codex receives the user's question.

2. Retrieve

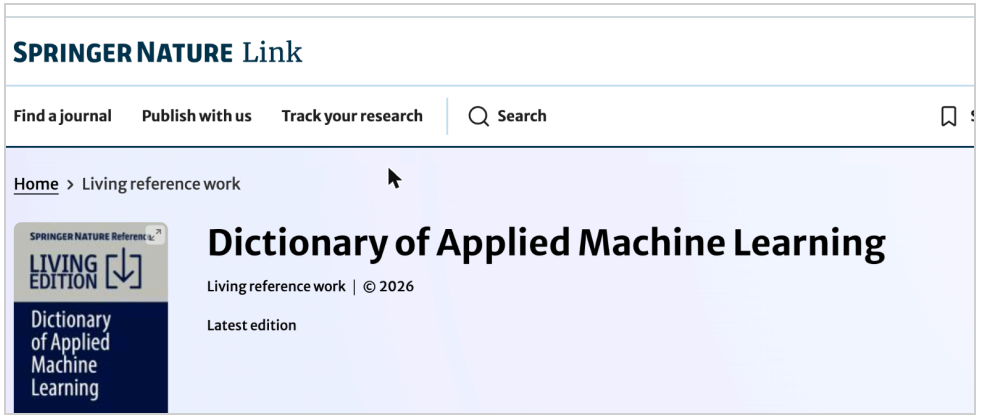
It calls **lookup_term** and **get_related_terms**.

3. Answer

The response is grounded in the published entry.

Open endpoint
dictionaryofml.org/mcp

The Springer reference work



Jung, A., Schnoor, E. and Olioumtsevits, K. (eds.) (2026). *Dictionary of Applied Machine Learning* (Living Reference Work). Singapore: Springer Nature Singapore. Online ISBN 978-981-92-2012-0. https://doi.org/10.1007/978-981-92-2012-0

Just released: the first Springer entries

Published online first on SpringerLink · 1 November 2026

data

Recorded information in a form a method can process: the data point, the dataset, the feature matrix.

loss

A scalar measuring the error of a hypothesis on one data point. Training minimizes its average.

regularization

Modifications that improve generalization: prune the hypothesis space, penalize the loss, augment the training set.

An audit can now cite the definition it relied on, and the version of it.

Jung, A. (2026). Data. In *Dictionary of Applied Machine Learning*. Springer Nature Singapore. https://doi.org/10.1007/978-981-92-2012-0_1-1

Jung, A. (2026). Loss. In *Dictionary of Applied Machine Learning*. Springer Nature Singapore. https://doi.org/10.1007/978-981-92-2012-0_3-1

Jung, A. (2026). Regularization. In *Dictionary of Applied Machine Learning*. Springer Nature Singapore. https://doi.org/10.1007/978-981-92-2012-0_8-1

From FLAIG research to shared infrastructure

Research

Trustworthiness concepts made precise for models trained from data.

Practice

A common language for developers, compliance teams and regulators.

Infrastructure

Definitions that people can read, run, query and cite.

Read · run · query · cite

dictionaryofml.org

Corrections and new terms welcome: alex.jung@aalto.fi