Datasheet for Dataset: Standardized Competency Catalog

Auto-generated by scripts/generate_datasheet.py

2026-02-16

Contents

1 Section A: High-Level Summary

1.1 Motivation

This datasheet documents the structure and coverage of the standardized competency catalog for machine-learning and analytics workflows.

1.2 Composition

- Input file: standardized-comptency-catalog.json
- Generated at: `2026-02-16 18:03:50`
- Number of knowledge areas: **17**
- Number of competencies: **208**
- Number of sources: **1**
- Average competencies per knowledge area: **12.24**
- Competencies with empty description: **0**

1.3 Distribution Snapshot (Table)

Knowledge Area	Competencies
Society, Ethics, and the Profession	21
Artificial Intelligence	15
Data Management	15
Operating Systems	15
Foundations of Programming Languages	14
Networking and Communication	14
Architecture and Organization	13
Parallel and Distributed Computing	13
Algorithmic Foundations	12
Graphics and Interactive Techniques	12

1.4 Distribution Snapshot (Chart)

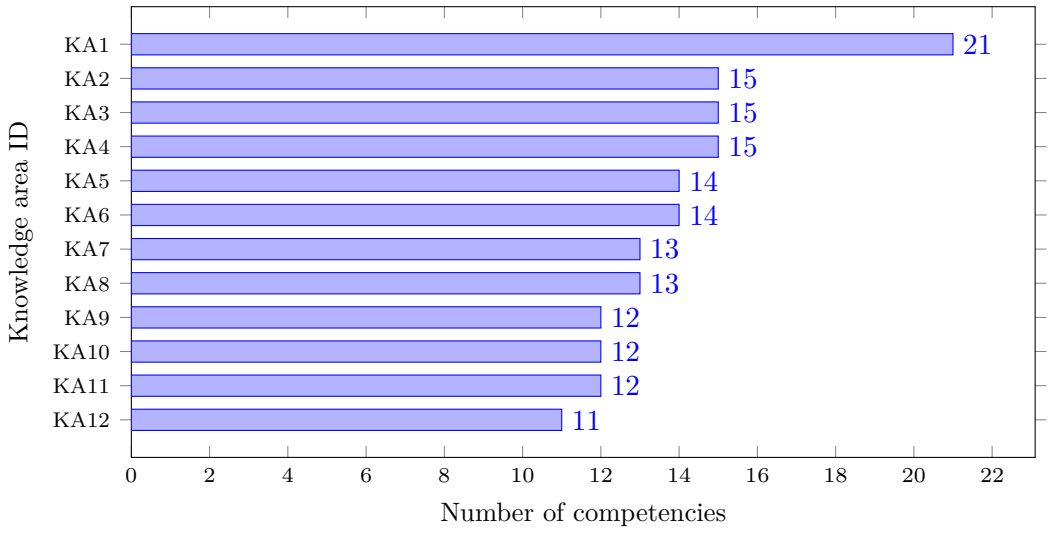

Figure 1: Competencies per knowledge area (top entries by count). See KA mapping table below.

ID	Knowledge Area
KA1	Society, Ethics, and the Profession

ID	Knowledge Area
KA2	Artificial Intelligence
KA3	Data Management
KA4	Operating Systems
KA5	Foundations of Programming Languages
KA6	Networking and Communication
KA7	Architecture and Organization
KA8	Parallel and Distributed Computing
KA9	Algorithmic Foundations
KA10	Graphics and Interactive Techniques
KA11	Security
KA12	Human-Computer Interaction

1.5 Quality Check Summary

Check	Result
Missing descriptions	0
Missing sourceId references	0
Invalid sourceId references	0
Duplicate knowledge area titles	0
Duplicate competency titles (global)	6

2 Section B: Detailed Datasheet (Datasheets for Datasets)

2.1 Collection Process

Collection process details are not explicitly encoded in this JSON dataset and should be documented by dataset maintainers in future revisions.

2.2 Preprocessing / Cleaning / Labeling

The dataset appears to be pre-structured by knowledge area with competency records containing title, description, taxonomy, sourceId, and version fields.

2.3 Uses

Potential uses include curriculum mapping, competency gap analysis, and as a structured reference for educational ML/NLP workflows. This datasheet does not validate downstream model suitability.

2.4 Distribution

This report was generated from a local JSON file and can be redistributed as PDF/Markdown/JSON outputs.

2.5 Maintenance

Recommended maintenance: version dataset snapshots, monitor schema drift, and re-run this generator in CI for each update.

2.6 Detailed Composition Tables

2.6.1 Competencies per knowledge area

Knowledge Area	Competencies
Society, Ethics, and the Profession	21
Artificial Intelligence	15
Data Management	15
Operating Systems	15
Foundations of Programming Languages	14
Networking and Communication	14
Architecture and Organization	13
Parallel and Distributed Computing	13
Algorithmic Foundations	12
Graphics and Interactive Techniques	12
Security	12
Human-Computer Interaction	11
Software Engineering	10
Specialized Platform Development	10
Systems Fundamentals	10
Mathematical and Statistical Foundations	6
Software Development Fundamentals	5

2.6.2 Taxonomy distribution

Taxonomy	Count
UNDERSTAND	98
CREATE	45
EVALUATE	34
APPLY	30
REMEMBER	1

2.6.3 Version distribution

Version	Count
1.0.0	208

2.6.4 Source ID usage

Source ID	Count
1	208

2.7 Intermediate Quality Checks

2.7.1 Schema sanity

Check	Value
Invalid knowledge area item types	0
Invalid competency item types	0

2.7.2 Missing fields

Field	Missing Count
(none found)	0

2.7.3 Duplicate detection

Check	Count
Duplicate knowledge area titles	0
Duplicate competency titles (global)	6

2.7.4 Invalid source references

Source ID	Invalid Reference Count
(none found)	0

2.8 Source Catalog

Source ID	Title	Author	URI
1	Computer Science Curricula 2023	The Joint Task Force on Computing Curricula	https://csed.acm.org/cs2023-report-with-feedback/

2.9 Limitations

This report is descriptive. It does not verify semantic correctness of competency descriptions or evaluate model performance impact.