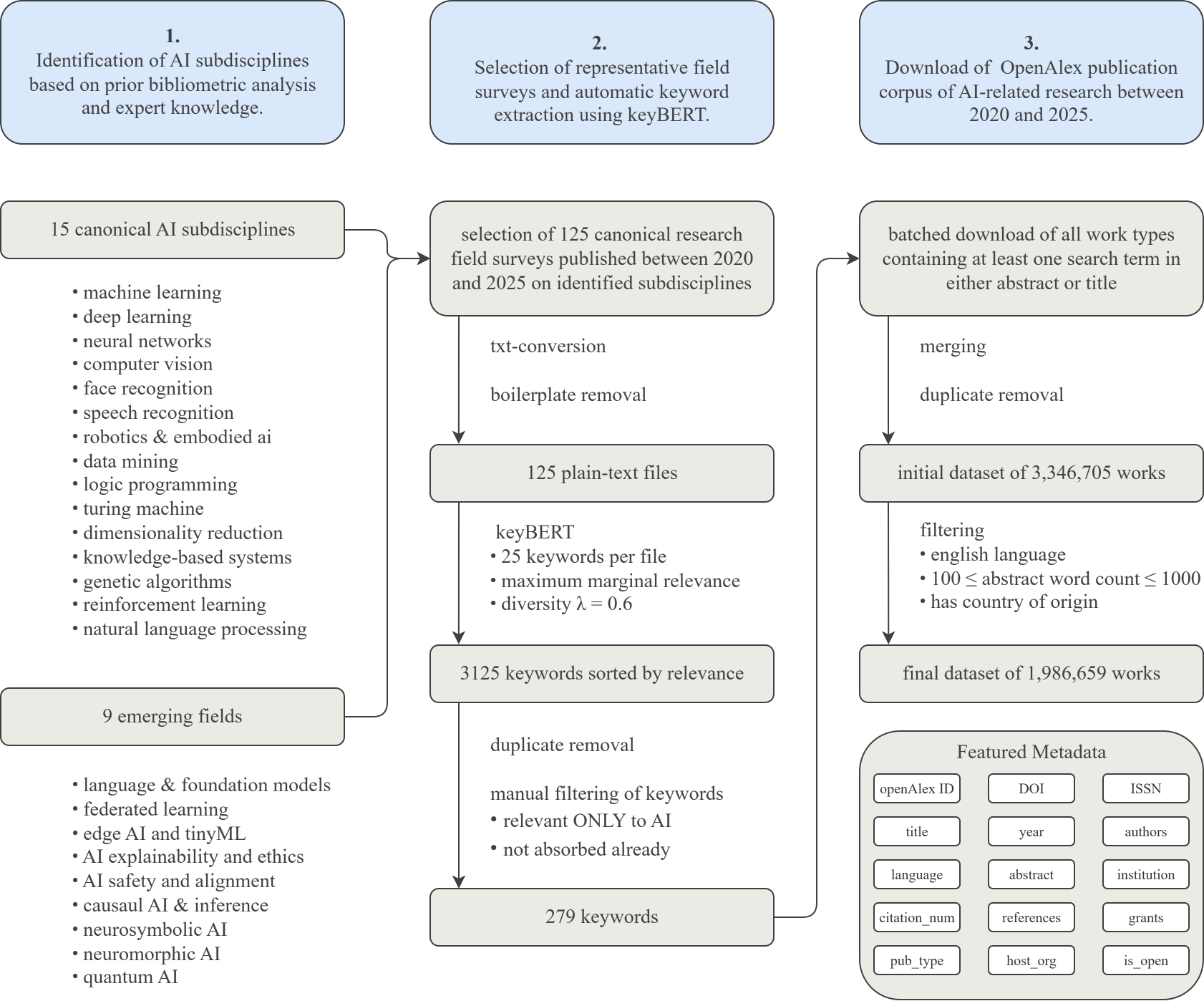
1.
Identification of AI subdisciplines based on prior bibliometric analysis and expert knowledge.
2.
Selection of representative field surveys and automatic keyword extraction using keyBERT.
3.
Download of OpenAlex publication corpus of AI-related research between 2020 and 2025.
15 canonical AI subdisciplines
• machine learning
• deep learning
• neural networks
• computer vision
• face recognition
• speech recognition
• robotics & embodied ai
• data mining
• logic programming
• turing machine
• dimensionality reduction
• knowledge-based systems
• genetic algorithms
• reinforcement learning
• natural language processing
9 emerging fields
• language & foundation models
• federated learning
• edge AI and tinyML
• AI explainability and ethics
• AI safety and alignment
• causaul AI & inference
• neurosymbolic AI
• neuromorphic AI
• quantum AI
selection of 125 canonical research field surveys published between 2020 and 2025 on identified subdisciplines
txt-conversion
boilerplate removal
125 plain-text files
keyBERT
• 25 keywords per file
• maximum marginal relevance
• diversity λ = 0.6
3125 keywords sorted by relevance
duplicate removal
manual filtering of keywords
• relevant ONLY to AI
• not absorbed already
279 keywords
batched download of all work types containing at least one search term in either abstract or title
merging
duplicate removal
initial dataset of 3,346,705 works
filtering
• english language
• 100 ≤ abstract word count ≤ 1000
• has country of origin
final dataset of 1,986,659 works
Featured Metadata
openAlex ID
DOI
ISSN
title
year
authors
language
abstract
institution
citation_num
references
grants
pub_type
host_org
is_open