MACHINE LEARNING FOR ABSTRACT SCREENING

A TUTORIAL ON HOW TO USE ASREVIEW



Daniel Ebbert · Adelaide University

23 July 2026

GOAL FOR TODAY

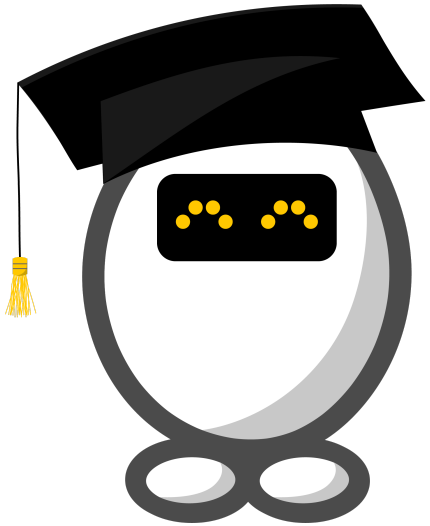
Enable you to save time during abstract screening for systematic reviews.



ASREVIEW

- Open-source tool developed by the **AI-aided Knowledge Discovery Lab** at Utrecht University
- Designed to streamline title and abstract screening in systematic reviews using machine learning

https://asreview.nl/



ABOUT THIS PRESENTATION

This presentation is based on the tutorial paper by Fütterer et al. (2026). The original work and recommendations are theirs; this workshop makes their guidance accessible for our group.

Fütterer, T., König, L., Campos, D. G., Scherer, R., Zitzmann, S., & Hecht, M. (2026). Fast-track your abstract screening: Mastering ASReview for accelerating abstract screening and evaluating decisions from automatic-screening methods. Advances in Methods and Practices in Psychological Science, 9(2), Article 25152459261442150. https://doi.org/10.1177/25152459261442150

THE PROBLEM

Systematic reviews are essential for evidence-based decision-making, but producing them is highly time-consuming.

- Screening a single title and abstract takes **~30 seconds** on average
- Screening **3,000 abstracts** takes roughly **3 work days**
- Traditional screening has an **~10% error rate**, often due to fatigue

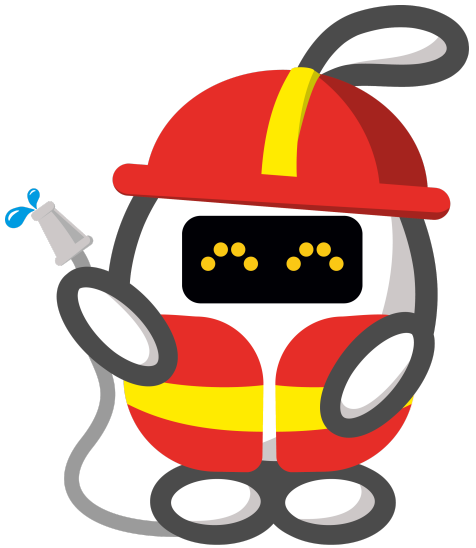


THE BOTTLENECK

Abstract screening is a labour-intensive stage of a systematic review.

The challenge is twofold:

- Identify as many relevant articles as possible (comprehensiveness)
- Keep the screening workload feasible (efficiency)

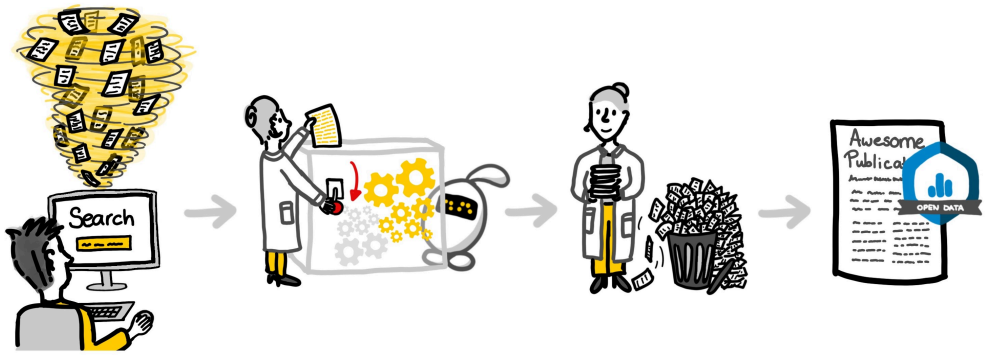


CAN MACHINE LEARNING HELP?

ML tools offer promise for efficiency, accuracy, and bias reduction.

One approach is **active learning**, where the model learns from screener decisions and progressively improves its predictions.

But many researchers are unfamiliar with them and unsure how to integrate them into their workflow.

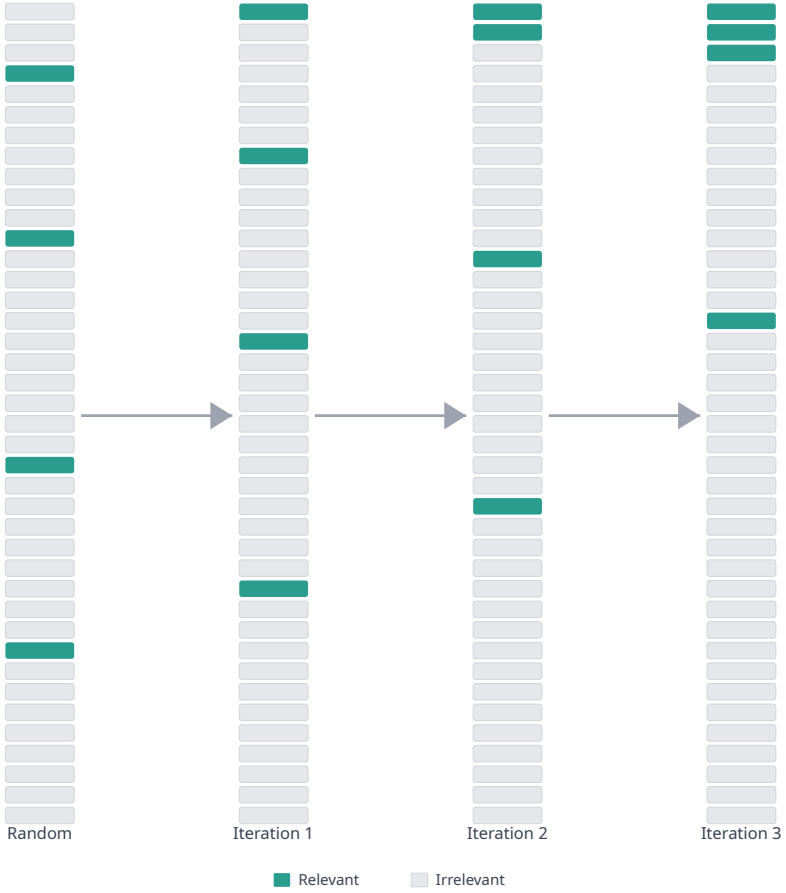


HOW ACTIVE LEARNING WORKS

1. The screener labels abstracts as relevant or irrelevant
2. These labels serve as training data for an ML model
3. The model ranks remaining abstracts by predicted relevance
4. The screener screens the next abstract (most likely relevant)
5. The cycle repeats, progressively improving predictions



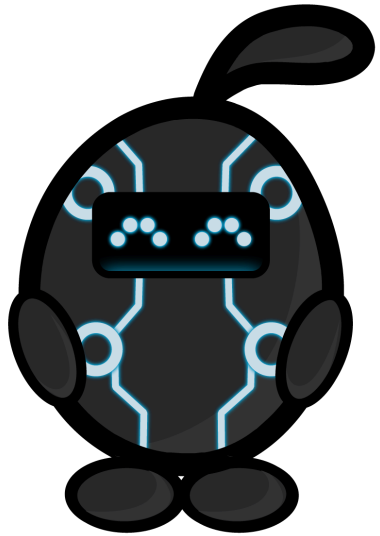
HOW ACTIVE LEARNING WORKS



KEY COMPONENTS OF THE ML ALGORITHM

The algorithm has two parts:

- **Feature extractor**: identifies linguistic elements to assess similarities between abstracts
  - TF-IDF: weights words by frequency and rarity across the corpus
  - SBERT: pretrained language model capturing sentence-level semantics
- **Classifier**: calculates relevance scores from the feature extractor's output
  - Options: Logistic Regression, SVM, Naive Bayes, Random Forest



Recommended combination by Fütterer et al. (2026): SBERT + Logistic Regression

Developer recommendation and default setting: TF-IDF + SVM

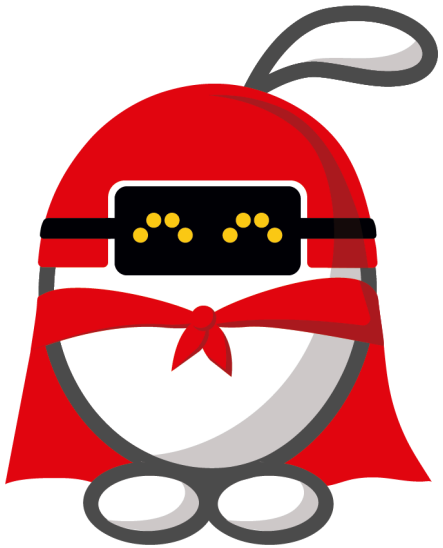
EFFICIENCY GAINS

An average of **30-39% of abstracts** need to be screened to identify 95% of relevant articles.

For a dataset of 3,000 abstracts (at ~30 seconds per abstract):

- **Manual screening:** ~25 hours, roughly 3 work days
- **With ASReview (35%):** ~9 hours, roughly 1 work day

The final 5% of relevant studies considerably increases screening time, which is why a 95% identification threshold is widely accepted.



WHEN TO USE ASREVIEW

ASReview is most valuable for large datasets.

- Up to ~1,000 studies: manual screening is usually feasible
- 7,000-40,000 abstracts: ASReview saves a tremendous amount of work
- 300-400 studies: probably not worth it



THE SCREENING WORKFLOW

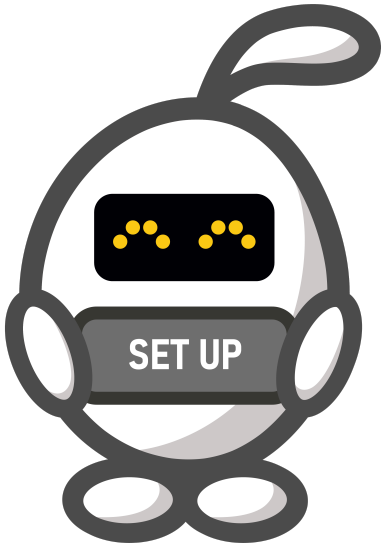
Three steps:

1. **Prescreening** (random)
2. **Stopping-rule selection**
3. **ML-aided screening**

STEP 1: PRESCREENING

Screen randomly until at least one relevant and one irrelevant abstract have been identified.

Screen at least **1% of the dataset** or **100 abstracts** (whichever is larger).



STEP 2: STOPPING RULES

A major challenge: determining when to stop screening.

Goal: identify approximately **95% of relevant literature**.

First, estimate prevalence from your prescreening sample:

Prevalence = % relevant during prescreening (e.g., 10 out of 100 = 10%)

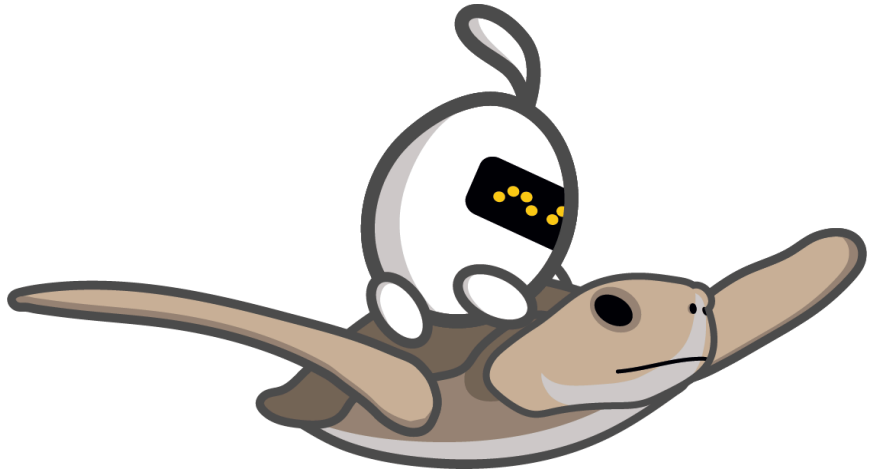
Prevalence determines which stopping rule to use.



STOPPING HEURISTICS

Two heuristics:

- **Time-based**: stop after X% of the dataset has been screened
- **Data-driven**: stop after X consecutive irrelevant abstracts



STOPPING RULES BY PREVALENCE

Prevalence determines which rules apply:

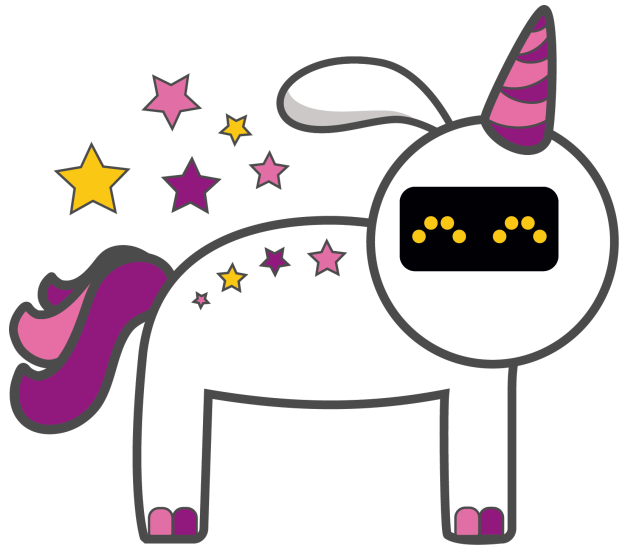
Prevalence	Rule	Time-based	Data-driven	Breakout
≤ 5%	A	35%	7%	15%
> 5% to 10%	B	25%	5%	10%
> 10%	C	15%	3%	10%

The breakout rule allows early stopping if a large consecutive run of irrelevant abstracts occurs.



STEP 3: ML-AIDED SCREENING

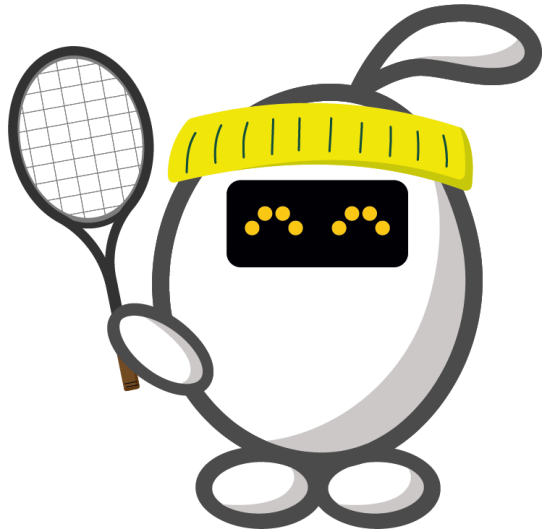
1. Create a new project (recommended) or continue in the same project
2. Configure model: SBERT + Logistic Regression + Maximum query
3. Screen abstracts in order of predicted relevance
4. ASReview auto-stops when the configured heuristic is triggered



MODEL CONFIGURATION

Use a custom configuration:

- **Feature extractor**: SBERT
- **Classifier**: Logistic Regression
- **Query strategy**: Maximum (certainty-based)



A NOTE ON PRIOR KNOWLEDGE

The tutorial recommends **not** using the prior knowledge function.

- The model needs **one relevant and one irrelevant** abstract for initial training
- When no prior knowledge is provided, screening begins randomly until both are found
- Adding more than one relevant study has little impact on performance
- Prior knowledge risks selection bias



PREREGISTRATION AND REPORTING

- Steps 1 and 2 can be **preregistered** before screening begins
- Preregistration can be updated before ML-aided screening to include specific stopping strategies
- Templates for preregistering and reporting are available on OSF
- Record: model details, stopping rules, (estimated) prevalence





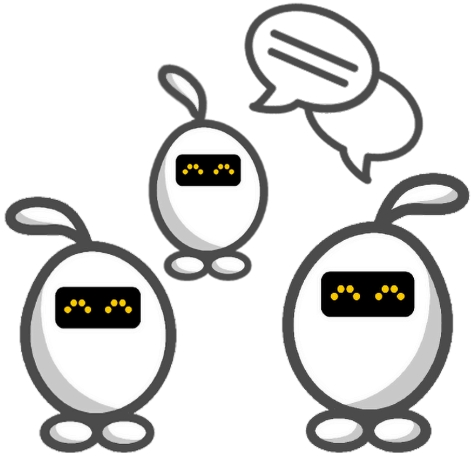
osf.io/xrb9z

COLLABORATIVE SETTINGS: THE CHALLENGE

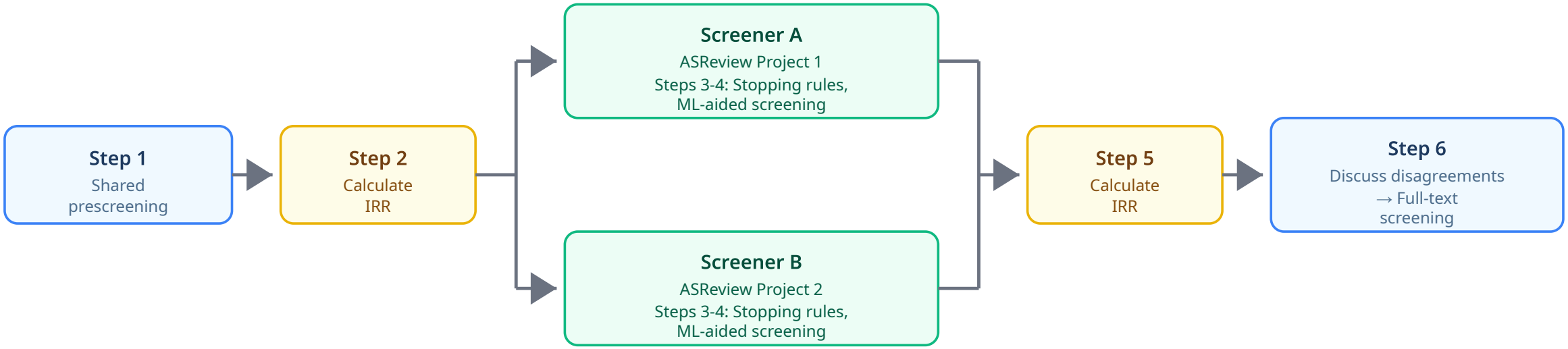
ASReview was **not originally designed** for collaborative double screening.

There is currently no standardised, evidence-based procedure for this scenario.

However, some teams implement double screening by creating **two separate ASReview projects**, each with the same or different data sets, training data, and learning algorithms, and then comparing results to identify disagreements.



A COLLABORATIVE WORKFLOW



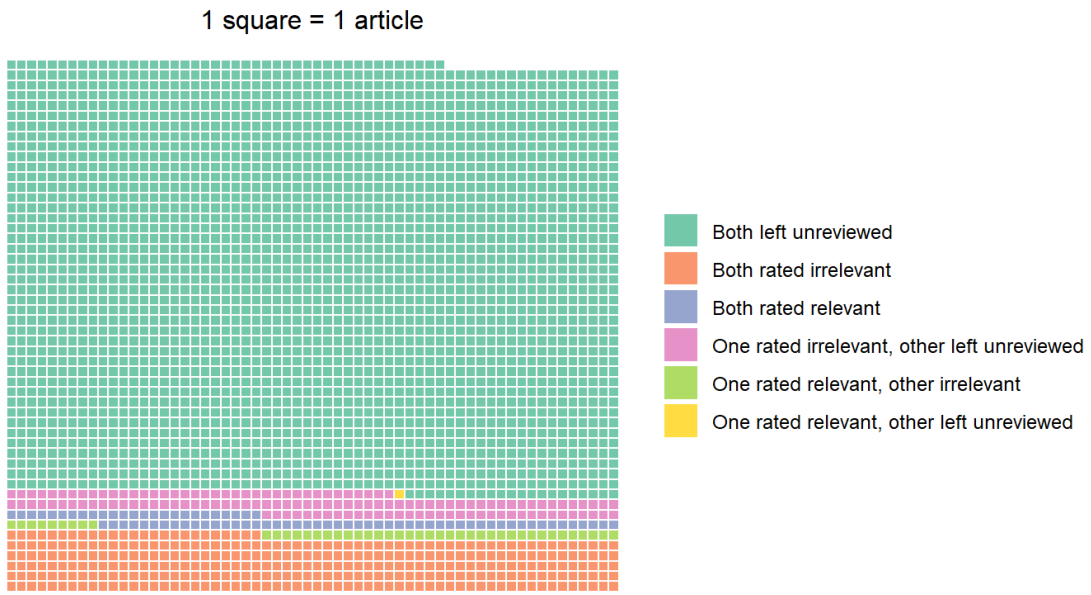
HOW TO DETERMINE IRR?

The Problem:

Standard inter-rater reliability (IRR) requires fully overlapping ratings.

The Solution:

Resolved by `asreview_irr`, an R-based tool that automatically parses exported ASReview data to calculate robust agreement metrics.



Repository:
github.com/langtonhugh/asreview_irr

SUMMARY

- Abstract screening is a major bottleneck in systematic reviews
- ASReview uses active learning to prioritise relevant abstracts, saving substantial time
- Follow three steps: prescreening, stopping-rule selection, ML-aided screening
- Use LR + SBERT, estimate prevalence, apply the appropriate stopping rule
- For collaborative use: two separate projects, establish IRR first
- Best for large datasets (1,000+); consider manual screening for smaller sets

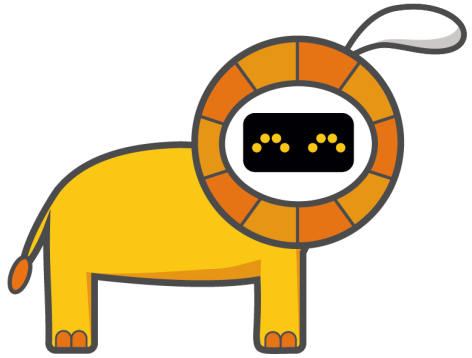


HANDS-ON SESSION

Use the C3L-hosted instance:

https://asreview.c3l.ai

No installation required. Just create an account and log in.



HANDS-ON: GET THE DATA

Download the practice dataset for this session:

http://bit.ly/4ffwVgW

This includes the raw RIS file ready to be imported into ASReview.

bit.ly/4ffwVgW

THANK YOU

Questions?

Daniel Ebbert
daniel.ebbert@adelaide.edu.au



SOURCES & ATTRIBUTIONS

TUTORIAL PAPER

Fütterer, T., König, L., Campos, D. G., Scherer, R., Zitzmann, S., & Hecht, M. (2026). Fast-track your abstract screening: Mastering ASReview for accelerating abstract screening and evaluating decisions from automatic-screening methods. *Advances in Methods and Practices in Psychological Science*, *9*(2), Article 25152459261442150. https://doi.org/10.1177/25152459261442150

ARTWORK

Mascot (Elas) and project illustrations by **Joukje Willemsen** (Licensed under CC BY-NC-ND 4.0).
Source repository: https://github.com/asreview/asreview-artwork

LICENSE

These slides are licensed under **Creative Commons Attribution 4.0 International (CC BY 4.0)**.

This licence does not cover the included artwork and illustrations, which are licensed under their own terms (CC BY-NC-ND 4.0).