Package 'aieconindex'

May 13, 2026

Title Access the 'Anthropic Economic Index' Dataset

Version 0.1.0

Description Provides clean, tidy access to the 'Anthropic Economic Index' (AEI) dataset hosted on 'Hugging Face' <https://huggingface.co/datasets/Anthropic/EconomicIndex>. The AEI is a recurring release from 'Anthropic' that maps usage of the 'Claude' family of large language models to occupations and tasks using the 'O*NET' taxonomy and the 'Standard Occupational Classification' system, following the methodology of Handa et al. (2025) <doi:10.48550/arXiv.2503.04761> and the privacy-preserving system 'Clio' of Tamkin et al. (2024) <doi:10.48550/arXiv.2412.13678>. Functions list available releases, fetch raw and enriched usage tables, retrieve task statements, request hierarchies, and country-level breakdowns, compare two releases, join the index to user-supplied data on a shared key, and compute usage-concentration metrics (Herfindahl-Hirschman Index, top-N concentration ratios, Shannon entropy). Data is cached locally for subsequent calls. Reproducibility helpers produce 'BibTeX' or plain-text citations that include the methodological source paper. This product uses the 'Anthropic Economic Index' data (released under CC-BY by 'Anthropic') but is not endorsed or certified by 'Anthropic'.

License MIT + file LICENSE

Encoding UTF-8

Language en-US

URL https://github.com/charlescoverdale/aieconindex

BugReports https://github.com/charlescoverdale/aieconindex/issues

RoxygenNote 7.3.3

Depends R (>= 4.1.0)

Imports cli (>= 3.6.0), httr2 (>= 1.0.0), jsonlite, stats, tools, utils

Suggests knitr, rmarkdown, testthat (>= 3.0.0), withr

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Charles Coverdale [aut, cre]

Maintainer Charles Coverdale <charlesfcoverdale@gmail.com>

Repository CRAN

Date/Publication 2026-05-13 20:10:02 UTC

Contents

`aei_cache_clear` *Clear the aieconindex cache*

Description

Deletes all locally cached AEI files. The next call to any data function will re-download from Hugging Face.

Usage

```
aei_cache_clear()
```

Value

Invisible `NULL`.

See Also

Other configuration: aei_cache_info()

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
aei_cache_clear()
options(op)
```

---

aei_cache_dir *Locate the aieconindex cache directory*

Description

Returns the directory used to store downloaded AEI files. Defaults to tools::R_user_dir("aieconindex", "cache"). Override by setting options(aieconindex.cache_dir = "/your/path").

Usage

```
aei_cache_dir()
```

Value

A character string giving the absolute path.

---

aei_cache_info *Inspect the local aieconindex cache*

Description

Returns information about the local cache: where it lives, how many files it contains, and how much disk space they take.

Usage

```
aei_cache_info()
```

Value

A list with elements dir, n_files, size_bytes, size_human, and files (a data frame with name, size_bytes, and modified columns).

See Also

Other configuration: aei_cache_clear()

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
aei_cache_info()
options(op)
```

`aei_cite` *Citation strings for the Anthropic Economic Index*

Description

Returns a citation for either the Anthropic Economic Index project as a whole or a specific release, in the requested format. The Anthropic Economic Index data is released under Creative Commons Attribution 4.0 International (CC-BY-4.0); attribution is required when redistributing the data.

Usage

```
aei_cite(
  release = "all",
  format = c("text", "bibtex", "bibentry"),
  method = TRUE
)
```

Arguments

`release`	A release identifier, or `"all"` (the default) to cite the project rather than a specific release.
`format`	One of `"text"`, `"bibtex"`, or `"bibentry"`.
`method`	Logical. If `TRUE` (the default), include the methodological source paper (Handa et al. 2025) in the BibTeX or bibentry output. Set to `FALSE` to return the dataset citation only.

Details

For `release = "all"` (the default), the citation refers to the methodological source paper: Handa, K. et al. (2025), "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations" (arXiv:2503.04761). For a specific release, the citation refers to the dataset snapshot at that release on Hugging Face, with the headline model and the Anthropic report PDF included when one is bundled.

Hugging Face datasets do not currently issue DOIs by default; the `url` field is the stable Hugging Face path. The methodological source paper is on arXiv and has a permanent identifier (arXiv:2503.04761).

Value

A character vector for `"text"` and `"bibtex"`; a `bibentry` object (possibly with multiple entries) for `"bibentry"`.

References

Handa, K., Tamkin, A., McCain, M., Huang, S., Durmus, E., Heck, S., Mueller, J., Hong, J., Ritchie, S., Belonax, T., Troy, K. K., Amodei, D., Kaplan, J., Clark, J., and Ganguli, D. (2025). Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations. arXiv:2503.04761. https://arxiv.org/abs/2503.04761

Tamkin, A. et al. (2024). Clio: Privacy-Preserving Insights into Real-World AI Use. arXiv:2412.13678. https://arxiv.org/abs/2412.13678

Examples

```
aei_cite()
aei_cite("2026-03-24", format = "bibtex")
aei_cite("2025-09-15", format = "bibentry")
aei_cite(format = "text", method = FALSE)
```

---

aei_clusters *Fetch the request-hierarchy tree for a release*

Description

The Anthropic Economic Index groups Claude requests into a multi-level hierarchy of clusters using the Clio privacy-preserving system. From the 2025-09-15 release onwards each release ships these trees as JSON files. This function fetches the relevant tree as a parsed nested list.

Usage

```
aei_clusters(
  release = "latest",
  source = c("claude_ai", "1p_api"),
  use_cache = TRUE
)
```

Arguments

release	A release identifier. See aei_files().
source	Character. Either "claude_ai" or "1p_api".
use_cache	Logical. If TRUE (the default), use the local cache when present.

Details

Clusters are produced by the Clio system (Tamkin et al. 2024), which summarises requests into facets, then groups them into a tree where each level represents a coarser grouping. The JSON returned has one entry per top-level cluster, each with name, optional description, optional count, and a list of children mirroring the same shape. Cluster summaries that fail Clio's privacy checks (low cell counts, identifying information) are dropped before the tree is published.

Releases before 2025-09-15 do not contain request-hierarchy trees.

Value

The parsed request hierarchy as a nested list.

References

Tamkin, A. et al. (2024). Clio: Privacy-Preserving Insights into Real-World AI Use. arXiv:2412.13678. https://arxiv.org/abs/2412.13678

See Also

Other core data: aei_download(), aei_geography(), aei_index(), aei_tasks()

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
tree <- aei_clusters("2025-09-15")
length(tree)
options(op)
```

---

aei_compare *Compare two Anthropic Economic Index releases*

Description

Side-by-side diff of the same metric across two releases. Useful for tracking how the share of conversations classified to a given O*NET task or country has shifted between two points in time.

Usage

```
aei_compare(
  release_a,
  release_b,
  source = c("claude_ai", "1p_api"),
  variant = c("raw", "enriched"),
  by = c("cluster_name", "facet", "variable"),
  value_col = "value",
  use_cache = TRUE
)
```

Arguments

release_a, release_b	Release identifiers. See aei_files(). release_a is treated as the baseline.
source	Character. Either "claude_ai" or "1p_api".
variant	Character. Either "raw" or "enriched".

by	Character vector of join keys. Default is `c("cluster_name", "facet", "variable")` for the long-format schema. Add `c("geo_id", "geography")` to compare geography rows.
value_col	Character. Name of the numeric column to compare. Default `"value"` for long-format AEI tables.
use_cache	Logical. If `TRUE` (the default), use the local cache when present.

Details

The function fetches both releases via `aei_index()`, inner-joins them on the columns in `by`, and returns one row per shared key with both values plus the absolute and percentage change. The default join keys (`cluster_name`, `facet`, `variable`) are the natural composite key of the long-format AEI schema introduced in the 2025-09-15 release. For comparisons that include geographic breakdowns add `geo_id` and `geography` to `by`.

Releases that ship in different schemas (the wide-format 2025-02-10 and 2025-03-27 releases vs the long-format 2025-09-15+) cannot be compared directly. Use `aei_download()` and a hand-written join in that case.

Pct-change is calculated as `(value_b - value_a) / value_a * 100` and is `NA` where `value_a` is zero.

Value

An aei_tbl with the join keys plus `value_a`, `value_b`, `delta` (= `value_b - value_a`), and `pct_change`.

References

Handa, K. et al. (2025). Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations. arXiv:2503.04761. https://arxiv.org/abs/2503.04761

See Also

Other analysis: `aei_concentration()`, `aei_link()`

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
diff <- aei_compare("2025-09-15", "2026-03-24")
head(diff[order(-abs(diff$delta)), ])
options(op)
```

`aei_concentration` *Usage concentration metrics for an Anthropic Economic Index slice*

Description

Computes Herfindahl-Hirschman Index (HHI), top-N concentration ratios (CR4 by default), and Shannon entropy on a vector of usage shares. Useful for asking "how concentrated is Claude usage across tasks / occupations / countries?" against the same data the AEI reports for percentage shares.

Usage

```
aei_concentration(x, share_col = NULL, group_cols = NULL, top_n = 4L)
```

Arguments

`x`	A `data.frame` (or aei_tbl) with a column of shares.
`share_col`	Character. Name of the share column. Defaults to `"value"` (the long-format AEI column name) if present, otherwise `"pct"`.
`group_cols`	Optional character vector of grouping columns. If supplied, returns one row of metrics per group.
`top_n`	Integer. The N for the CR_n top-share metric. Default 4.

Details

Three measures are produced for each call:

- **HHI** = sum of squared shares. When shares are in percentages (0 to 100), HHI ranges from 0 (perfect dispersion) to 10,000 (one item holds all). When shares are in proportions (0 to 1), HHI ranges from 0 to 1. The function detects the scale automatically: if `max(share) > 1` the shares are treated as percentages, otherwise as proportions.
- **CR_n** = sum of the top-n shares. Defaults to CR4. Same units as the input.
- **Entropy** = Shannon entropy in bits, computed on the normalised proportions. Maximum entropy at uniform distribution is `log2(n)` where `n` is the number of non-zero shares.

Rows with `NA`, zero, or negative shares are dropped before computation. If a `group_cols` argument is supplied, the metrics are computed within each group.

Value

A `data.frame` with columns `n` (number of non-zero shares), `hhi`, `cr_n` (named after `top_n`, e.g. `cr_4`), `entropy_bits`, `entropy_max_bits` (= `log2(n)`), and `entropy_normalised` (`entropy_bits / entropy_max_bits`, on the unit interval). When `group_cols` is supplied, the grouping columns are prepended.

References

Hirschman, A. O. (1964). "The Paternity of an Index". *The American Economic Review*, 54(5), 761-762.

Shannon, C. E. (1948). "A Mathematical Theory of Communication". *Bell System Technical Journal*, 27(3), 379-423. doi:10.1002/j.15387305.1948.tb01338.x

See Also

Other analysis: `aei_compare()`, `aei_link()`

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
uk <- aei_geography("2025-09-15", country = "GBR")
uk_tasks <- uk[uk$facet == "onet_task" & uk$variable == "onet_task_pct", ]
aei_concentration(uk_tasks)
options(op)
```

`aei_download` *Download an arbitrary file from an Anthropic Economic Index release*

Description

Lower-level fetcher than `aei_index()`. Pass any path returned by `aei_files()` and get the file back as a parsed data frame (CSV) or list (JSON), or as a local path if the file extension is unrecognised.

Usage

```
aei_download(release, path, use_cache = TRUE)
```

Arguments

`release`	A release identifier. See `aei_files()`.
`path`	Character. A relative path within the release directory, for example `"data/aei_raw_claude_ai_2026-02`
`use_cache`	Logical. If `TRUE` (the default), use the local cache when present.

Details

CSV files are read with `utils::read.csv(stringsAsFactors = FALSE, check.names = FALSE)`, which preserves the original column names verbatim (the AEI uses dots and double colons in some column names that R would otherwise mangle). JSON files are parsed with `jsonlite::fromJSON(simplifyVector = FALSE)` to preserve the nested tree structure of the request hierarchies. For other extensions (e.g. PDF reports, PNG figures, IPython notebooks) the function returns the local cached path so that the caller can do whatever they like with the file.

Value

For CSV files, an aei_tbl. For JSON files, the parsed list. For other extensions, the absolute local path of the cached file.

See Also

Other core data: `aei_clusters()`, `aei_geography()`, `aei_index()`, `aei_tasks()`

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
aei_download("2025-03-27", "task_pct_v2.csv")
options(op)
```

`aei_files` *List files in an Anthropic Economic Index release*

Description

Returns the file tree for a single release directory on Hugging Face, descending into subdirectories. Useful for inspecting what raw files are available before calling `aei_download()` or `aei_index()`.

Usage

```
aei_files(release = "latest", recursive = TRUE)
```

Arguments

`release`	A release identifier. Either `"latest"`, a release id such as `"release_2026_03_24"`, or a date string `"2026-03-24"`.
`recursive`	Logical. If `TRUE` (the default), recurse into subdirectories. If `FALSE`, list only the top level of the release.

Value

An aei_tbl with columns `path`, `type`, and `size_bytes`.

See Also

Other release discovery: `aei_releases()`

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
aei_files("2026-03-24", recursive = FALSE)
options(op)
```

aei_geography *Filter the enriched usage table to country or US-state rows*

Description

From the 2025-09-15 release onward the Anthropic Economic Index ships a single long-format enriched CSV with one row per geography-facet-variable combination. Geographic breakdowns are rows in that table where the `geography` column is `"country"` or `"state_us"`. This function fetches the enriched table via `aei_index()` with `variant = "enriched"` and filters those rows.

Usage

```
aei_geography(
  release = "2025-09-15",
  source = c("claude_ai", "1p_api"),
  geography = c("country", "state_us"),
  country = NULL,
  use_cache = TRUE
)
```

Arguments

`release`	A release identifier. See `aei_files()`.
`source`	Character. Either `"claude_ai"` or `"1p_api"`. The 1P API release ships only `"global"` rows, so country filtering will typically return nothing for that source.
`geography`	Character. Either `"country"` or `"state_us"`.
`country`	Optional ISO 3166-1 alpha-3 country code in the enriched data (for example `"GBR"`, `"AUS"`, `"USA"`). If `NULL` (the default), all countries are returned.
`use_cache`	Logical. If `TRUE` (the default), use the local cache when present.

Details

The enriched table has columns `geo_id` (ISO-3 country code or US state code after enrichment), `geography` (one of `"country"`, `"state_us"`, `"global"`), `facet`, `variable`, `cluster_name`, and `value`. Setting `country = "GBR"` or `country = "AUS"` filters to that single country; the codes are ISO-3 in the enriched data. Setting `geography = "state_us"` returns the US-state breakdown instead of the country breakdown.

Releases before 2025-09-15 do not contain geographic data; calling `aei_geography()` on them returns an informative error.

Value

An aei_tbl containing the long-format geographic rows of the enriched usage table.

References

Handa, K. et al. (2025). Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations. arXiv:2503.04761. https://arxiv.org/abs/2503.04761

See Also

Other core data: aei_clusters(), aei_download(), aei_index(), aei_tasks()

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
uk <- aei_geography("2025-09-15", country = "GBR")
head(uk)
options(op)
```

---

aei_index *Fetch the main usage table for an Anthropic Economic Index release*

Description

Convenience wrapper that locates the canonical usage CSV for a release and returns it as a tidy data frame. The shape and exact filename of the canonical table varies across releases (the AEI refactored its directory layout in late 2025); this function papers over that variation by matching against well-known filename patterns.

Usage

```
aei_index(
  release = "latest",
  source = c("claude_ai", "1p_api"),
  variant = c("raw", "enriched"),
  use_cache = TRUE
)
```

Arguments

release	A release identifier. See aei_files() for the list of valid forms.
source	Character. Either "claude_ai" (Claude.ai consumer product traffic) or "1p_api" (first-party API traffic). Not all releases include both.
variant	Character. Either "raw" (counts and percentages straight from Anthropic's pipeline) or "enriched" (joined to O*NET / SOC metadata, with derived per-capita and tier metrics). Older releases may only ship one variant.
use_cache	Logical. If TRUE (the default), use the local cache when present. If FALSE, force a fresh download.

Details

File discovery uses the regular expression `aei_<variant>_<source>.*\.csv$` against the recursive file listing for a release. When more than one match exists (because a release may ship multiple date windows or revisions), the matches are sorted lexicographically descending and the first is used. Because the AEI uses ISO dates in filenames (e.g. `_2026-02-05_to_2026-02-12`), lexicographic sort approximates "most recent date window" but is not guaranteed to be correct if Anthropic changes its filename convention. Use `aei_files()` to inspect available files for a release if the heuristic surprises you, then use `aei_download()` to fetch a specific path.

Schema differs across releases:

- Releases up to and including 2025-03-27 ship wide-format tables (one row per occupation/task, columns for shares).
- Releases from 2025-09-15 onward ship long-format tables (one row per geography-facet-variable combination, with a single `value` column).

See `data_documentation.md` in each release directory for the authoritative schema.

Value

An aei_tbl containing the usage table.

References

Handa, K. et al. (2025). Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations. arXiv:2503.04761. https://arxiv.org/abs/2503.04761

See Also

Other core data: `aei_clusters()`, `aei_download()`, `aei_geography()`, `aei_tasks()`

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
aei_index("2025-09-15", variant = "enriched")
options(op)
```

`aei_link` *Join an Anthropic Economic Index table to your own data*

Description

Generic merge helper that preserves the aei_tbl class and provenance metadata. Use it to splice the AEI to any external data frame on a shared key column, for example joining country-level AEI shares to working-age employment counts from a national labour-force survey, or joining O*NET task identifiers to a user-supplied occupational crosswalk (SOC to ANZSCO, SOC to ISCO, SOC to SOC2020 UK, etc.).

Usage

```
aei_link(
  x,
  y,
  by,
  type = c("left", "inner", "full"),
  suffixes = c(".aei", ".y")
)
```

Arguments

x	An aei_tbl returned by any data-fetching function.
y	A data.frame (or aei_tbl) to join.
by	Character vector of column names present in both x and y. Pass a named vector (e.g. c(cluster_name = "onet_id")) to join columns with different names.
type	One of "left" (the default; keep all rows of x), "inner" (keep only matched rows), or "full" (keep all rows of both, fill with NA).
suffixes	Character vector of length two giving suffixes to append to overlapping column names.

Details

The function is a thin wrapper over base::merge() with two differences. First, it preserves the aei_tbl class and the aei_query provenance attribute on the returned object so that downstream code can still see where the AEI side of the join came from. Second, it warns when a join produces zero rows, which is usually a sign of a key mismatch (typed differently, different code system, or different case).

For occupational crosswalks: the long-format AEI schema (from 2025-09-15 onwards) carries the O*NET task identifier in the cluster_name column when facet == "onet_task", and SOC major group codes appear in cluster_name when facet == "onet_task" and variable == "soc_pct". See data_documentation.md in each release on Hugging Face for the authoritative schema.

For country joins: country codes are ISO-3 in the enriched data ("GBR", "AUS", "USA"). If your external data uses ISO-2 codes, map them first with a small lookup table or with the **countrycode** package on CRAN.

Value

An aei_tbl with the joined columns. Provenance metadata from x is preserved.

See Also

Other analysis: aei_compare(), aei_concentration()

Examples

```
# Join AEI country shares to a small external table of GDP per capita
country <- aei_geography("2025-09-15")
overlay <- data.frame(
  geo_id = c("GBR", "AUS", "USA"),
  gdp_pc = c(48000, 65000, 80000)
)
joined <- aei_link(country, overlay, by = "geo_id")
head(joined)
```

---

aei_releases *List available Anthropic Economic Index releases*

Description

Queries the Hugging Face dataset listing for the Anthropic Economic Index and returns one row per release, augmented with the headline Claude model and a short note when the release is recognised. When the network is unavailable (or `live = FALSE`), the function returns the bundled list of releases known at package build time.

Usage

```
aei_releases(live = TRUE)
```

Arguments

`live`	Logical. If `TRUE` (the default), query Hugging Face for the current set of release directories and merge with the bundled metadata. If `FALSE`, return the bundled list only.

Value

An aei_tbl with columns `release_id`, `release_date`, `model`, and `notes`.

See Also

Other release discovery: `aei_files()`

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
aei_releases()
aei_releases(live = FALSE)
options(op)
```

aei_tasks *Fetch the O*NET task statements bundled with a release*

Description

Returns the table of O*NET task statements that the Anthropic Economic Index uses as its task taxonomy.

Usage

```
aei_tasks(release = "2025-03-27", use_cache = TRUE)
```

Arguments

release	A release identifier. See aei_files().
use_cache	Logical. If TRUE (the default), use the local cache when present.

Details

The O*NET task statements file (onet_task_statements.csv) is shipped alongside the 2025-03-27 release; later releases reference O*NET through the enriched index file rather than redistributing the statements separately. The default release argument is set to "2025-03-27" for that reason. For later releases, the same task identifiers can be joined back from aei_index(release, variant = "enriched") where the cluster_name column carries the O*NET task identifier when facet == "onet_task".

Value

An aei_tbl containing the task statements.

References

U.S. Department of Labor, Employment and Training Administration. O*NET Database. https://www.onetonline.org/

Handa, K. et al. (2025). Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations. arXiv:2503.04761. https://arxiv.org/abs/2503.04761

See Also

Other core data: aei_clusters(), aei_download(), aei_geography(), aei_index()

Examples

```
op <- options(aieconindex.cache_dir = tempdir())
aei_tasks("2025-03-27")
options(op)
```

aei_tbl *The aei_tbl class*

Description

An `aei_tbl` is a `data.frame` returned by all data-fetching functions in this package. It carries provenance metadata as the `aei_query` attribute, and dispatches a custom `print()`, `summary()`, and `[` method that preserves the metadata when the table is subset.

Details

Inspect the metadata directly with `attr(x, "aei_query")`.

Value

An object of class `aei_tbl`, which inherits from `data.frame`.

Examples

```
df <- data.frame(a = 1:3)
attr(df, "aei_query") <- list(endpoint = "demo", release = "rel")
class(df) <- c("aei_tbl", "data.frame")
print(df)
```

print.aei_tbl *Print method for aei_tbl*

Description

Prepends a one-line provenance header summarising the query.

Usage

```
## S3 method for class 'aei_tbl'
print(x, ...)
```

Arguments

`x`	An `aei_tbl`.
`...`	Passed to the underlying `print.data.frame` method.

Value

`x`, invisibly.

summary.aei_tbl *Summary method for aei_tbl*

Description

Summary method for aei_tbl

Usage

```
## S3 method for class 'aei_tbl'
summary(object, ...)
```

Arguments

object	An aei_tbl.
...	Passed to the underlying summary.data.frame method.

Value

Invisibly returns the standard data frame summary.

[.aei_tbl *Subset method for aei_tbl*

Description

Preserves the aei_tbl class and aei_query attribute when subsetting.

Usage

```
## S3 method for class 'aei_tbl'
x[i, j, ..., drop = TRUE]
```

Arguments

x	An aei_tbl.
i	Row selector.
j	Column selector.
...	Other arguments passed to [.data.frame.
drop	Logical. As in [.data.frame.

Value

An aei_tbl (or a vector if drop collapses the result).

Index